

ri Feb 13 10:59:37 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus_p2n model

February 13, 2004, 01:35:16 ; Search time 7307 Seconds

on: (without alignments)
5100.406 Million cell updates/sec

US-09-812-350-17

fact score: 4550

quence: 1 MNPKFHKWETATATAHEL.....VKMKRIEIEDDDNEEMED 911

ring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

arched: 2889711 seqs, 20454813386 residues

al number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xl
=/cgn2_1/USPTO spool p/US09812350/runat_10022004_101515_13351/app query.fasta_1.1095
DB=genEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOPCL=0 -LOOPEXT=0
-LIST=45
-START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -ALIGN=15 -MODE=LOCAL
-THR_MIN=100 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-THR_SCORE=200 -THR_SCORE=200 -THR_SCORE=200 -THR_SCORE=200 -THR_SCORE=200
-HEAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
-OUTFMT=ptc -NEG SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG
-USER=US09812350 @CNC 1.1 3983 @runat_10022004_101515_13351 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6
-FGAPOP=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

atabase :

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.cm.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.ov.*
- 22: em.or.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

RESULT 1

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rtd.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4550	100.0	3105	8	ATU13949	UI3949 Arabidopsis
2	4545	99.9	2736	6	AX505540	AX505540 Sequence
3	4545	99.9	2736	6	AX589851	AX589851 Sequence
4	4388	96.4	6376	8	AF218796	AF218796 Arabidops
5	4388	96.4	5821	8	AC020579	AC020579 Arabidops
6	4003	88.0	3049	8	SOY8B100HS	L35272 Glycine max
7	3896	85.6	3052	8	AF083343	AF083343 Nicotiana
8	3883.5	85.4	3084	8	AF133840	AF133840 Zea mays
9	3840.5	84.4	3148	8	AF133281	AF133281 Oryza sat
10	3825	84.1	3107	8	AF097363	AF097363 Triticum
11	3822.5	84.0	3058	8	AF174433	AF174433 Triticum
12	3790	83.3	2821	8	AF083344	AF083344 Triticum
13	3691	81.1	145149	2	AC130611	AC130611 Oryza sat
14	3691	81.1	162429	2	AC132451	AC132451 Oryza sat
15	3658	80.4	3109	8	OSA316025	AF077337 Zea mays
16	3657	80.4	3942	8	AF077337	AF077337 Zea mays
17	2768	60.8	197976	8	ATCHRIV39	ATCHRIV39 Arabidops
18	2768	60.8	206606	8	ATFCA1	297336 Arabidopsis
19	2502.5	55.0	2045	8	AF083327	AF083327 Zea mays
20	2313	50.8	153750	2	AC129010	AC129010 Leishmani
21	2299.5	50.5	300300	1	AP005373	AP005373 Thermosyn
22	2298	50.5	7726	3	LDHSP100	294053 L. donovani
23	2294	50.4	10744	1	AE009461	AE009461 Brucella
24	2293	50.4	3179	1	BSU251205	BSU251205 Brucella
25	2293	50.4	10473	1	AE014477	AE014477 Brucella
26	2289	50.3	5997	3	LMCLPB	Z38058 L. major (MH
27	2288	50.3	145709	1	D90914	D90914 Synchocyst
28	2286	50.2	1570	8	AY062596	AY062596 Arabidops
29	2282	50.2	298425	1	AP005049	AP005049 Streptomy
30	2264	49.8	342800	1	AP003598	AP003598 Nostoc sp
31	2258.5	49.6	10382	1	AE007611	AE007611 Clostridi
32	2245.5	49.4	43980	6	AX413014	AX413014 Sequence
33	2245.5	49.4	258650	1	AL596171	AL596171 Listeria
34	2245.5	49.4	349980	6	AX417047	AX417047 Sequence
35	2243.5	49.3	303671	1	AE015937	AE015937 Clostridi
36	2242.5	49.3	113316	2	AC007866	AC007866 Trypanoso
37	2242.5	49.3	159717	2	AC008368	AC008368 Trypanoso
38	2241	49.3	304708	1	AE017001	AE017001 Bacilli
39	2238	49.2	301150	1	AP003190	AP003190 Clostridi
40	2236.5	49.2	295050	1	AL591982	AL591982 Listeria
41	2236.5	49.2	349980	6	AX641671	AX641671 Sequence
42	2236	49.1	300950	1	AP005940	AP005940 Bradyrhiz
43	2234.5	49.1	38644	8	SPBCL6D10	AL035637 S.pombe c
44	2230	49.0	302050	1	AL935257	AL935257 Lactobaci
45	2228	49.0	290029	1	AE017027	AE017027 Bacillus

ALIGNMENTS

Fri Feb 13 10:59:37 2004



ATU13949 3105 bp mRNA linear PLN 06-JUN-1995
Arabidopsis thaliana heat shock protein ATHSP101 (Athsp101) mRNA,
complete cds.
U13949
CDSSTON
RSION
YWORDS
URCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 3105)
Schirmer, B.C., Lindquist, S. and Vierling, E.
An Arabidopsis heat shock protein complements a thermotolerance
defect in yeast
Plant Cell 6 (12), 1899-1909 (1994)
95170291
7866032
2 (bases 1 to 3105)
Vierling, E.
Direct Submission
Submitted (24-AUG-1994) Elizabeth Vierling, Biochemistry,
University of Arizona, Life Sciences South, Tucson, AZ 85721, USA
Location/Qualifiers
1. 3105
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="1-18"
/tissue_type="leaf"
/clone_lib="lambda ZAP (Stratagene) cDNA library from heat
shock poly(A) RNA; K.W. Helm and E. Vierling, Nuc. Acids
Res. 17:7995 (1989)"
1. 3105
/gene="Athsp101"
/note="this protein is a member of the highly conserved
HSP100/Clp family of proteins, and like other HSP100/Clp
proteins, ATHSP101 has two consensus ATP-binding domains"
/codon_start=1
/product="ATHSP101"
/protein_id="AAA67927.1"
/db_xref="GI:537446"
/translation="MPEKFTHTNETIATAHELVNAGHAQFTPLHLAGALISDPTG
IPFAISSAGGENAASERVINQALKLPSPDPDDIPASSSLIKVIRRAQAOKS
RGDTLAVQLIMGLLEDSQIDLLNEGVATARKSEVEKLRGKGVESAGDTN
FQALTYGRDLVPAQGLDPFVIGRDEIRRVVRLSRRTKNPNPVLIGFVGKTAIVE
GLAQRIVGVNPSLTVRLISIDMGALVAGKRGFEERLKSVLKEVEDAEKVL
FIDEHLVGLAGTGMADANFKPLARGLQRCIATTEYKRYEKDAFFRR
QYVVAESPVDPTISILRGKLEKVEGHGVRIQDRLANAAQLSARYITGRHLPKAI
DLVDEACNRVQLDSOPEIDNLEKRMOLTELHALEREKDKASKALIEVRKLD
DLRDKLQPLTKYRKERIDEIRRLKQREELMFSIQEARRYDLARAADRLRYGAIQ
EVEISAIALEGTSESEENMUTENGVPEHIAEVSVRWITGIPVTLGQNEKRLIGLADR
LHKRVQGNQANVASAIIIRAGRAQQTGSFLFGTGVGKGTTELAKALAEOLF
DDENLVIRIDMSEVMEQSVSLIAGPGYVGEHGGQITVARRPYCVILFDEVEK
AHVAFNTLQVLDGDLQDQGTVDFRNSVIMTSLNLAHLLAGLTKGVMEVAR
DCVMREVRKHPRELLNDEIVDFDLSDQLRKVARLQMDVAVRLAERGVALAVT
DAALDYILAESIDPVGARPIRRMWEKVTLSKMWVREIDENSTVYIDAGAGLV
YRVESGLVDASTGKSDVLIHANGPKRSDAAQVKKMREIEIEDDNEEMIED"
2899...3105
/gene="Athsp101"
3105
/gene="Athsp101"
/note="19 A nucleotides"
891 a 541 c 852 g 821 t

BASE COUNT

ORIGIN

Alignment Scores: 7.12e-197 Length: 3105
Pred. No.: 4550.00 Matches: 911
Score: 100.00% Conservativity: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 8 Gaps: 0
DB: 8

US-09-812-350-17 (1-911) x ATU13949 (1-3105)

QY 1 MetAenProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 163 ATGAATCCAGAGAAATTCACACACAGACAAACGACAAATTCCTACAGCTCATGAGCTA 222
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
DB 223 GCTGTGAATCCAGGACATGCTCAATTCATCTCTCTTGCATTTAGCTGGTCTTGAATCTCT 282
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
DB 283 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGCGAGACGACGCTCAA 342
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
DB 343 TCTGCTGAAGAGTGAATCAATCAAGCCTTGAAGAAGTCTCTTCAATCTCTCCACCT 402
QY 81 AspAspIleProAlaSerSerSerIleLysValIleArgArgAlaGlnAlaAlaGln 100
DB 403 GATGATATTCACGCGAGTCTTAGTCTTATTAAGGTCTATTCGCTGGTCTCAAGCTCTAG 462
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
DB 463 AAGTCACGAGTGATATCTCATTTGGCTGTGTGACCATGATGATATGCGTCTCTTGAAGAT 522
QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
DB 523 TCTCAATCAGGAGTTGTTGAAGAGTGGTGTAGCGACGCGAGGCTAAAGCTCTGAG 582
QY 141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
DB 583 GTTCAGAGCTTCGTGGGAAAGAGGAGAAAGTGTAGAGTCTTCCGAGGAGAACGAG 642
QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
DB 643 TTTCAAGCTTTAAAGACTTATGAGAGAGATTTGGTTGAGCAGCAGGAGCTTATGATCT 702
QY 181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200
DB 703 GTGATTGCTGCTGATGAGAGATTAGAAGAGTCTGTAGAGTCTTCCGAGGAGAACGAG 762
QY 201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
DB 763 AACATCTCTGTCTTATTTGGAGAGCCAGGAGTGTGTAAACAGCTGTGTGAGGTTTA 822
QY 221 AlaGlnAlaGlyValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
DB 823 GCAAAAGAGTTGTGAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTCG 882
QY 241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu 260
DB 883 TTGGACATCGGTGCTTATTTGCTGTCTTAAATCCGAGGAGTGTGAAGAGGTTG 942
QY 261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
DB 943 AAATCTGTTTTGAAGAAGTTGAGGACCTGAGGCAAGTGAATCTCTTCTTATTGATGAG 1002
QY 281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
DB 1003 ATTCAATTTGTTCTTGTGCTGTGCAAACTGAAGGCTGATGAGTGAATCTGTTTC 1062
QY 301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
DB 1063 AAGCCCATGTTAGTAGAGGCGAGCTTCATGATCATTTGGTGTACACGCTTGAAGATAC 1122

321 ArgLysTyrValGluLysAspAlaAlaPheGluArgPheGlnGlnValTyrValAla 340
 1123 AGGAAATATGTGAGAAAGATGCTGCTTTGAGAGGAGGTTCACAAAGTCTAIGTTGG 1182
 341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
 1183 GAGCAAGTGTGCTGACACCAITAGTATCTTAGAGGACTCAAGGAGAGTATGAGGGA 1242
 361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
 1243 CATCATGCTGTGCGAATCCCAAGACAGAGCTTTATATAATGCTGCTCAGCTGTCTGCTGT 1302
 381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
 1303 TACATACTGTGCTGCGAATTTACCGGATTAAGCAATTTGATTTGGTTGATGAGGCTTGTGG 1362
 401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
 1363 AATGTGAGAGTCCAGCTTGATGATCAACCTCAAGAGATTGATAACCTTGAAGCAAGAGG 1422
 421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
 1423 ATGCAGCTGGAAATTTGAACCTTCAGCCCTTGGAAAGGAGAGGATAAAGCCAGCAAGCT 1482
 441 ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr 460
 1483 CGACTTATAGAGGTTCGGGAAGAGCTTGATGACCTGAGAGCAAGCTTTCAGCCCTCTCAG 1542
 461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
 1543 ATGAAATACAAAGAGGAGAGAGAGATTTGATGAGATTGCAAGGCTTAAACAGAAAGA 1602
 481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAla 500
 1603 GAAGAGCTCATGTTTCTTTCAGAGGAGCAGAGAGATATGACCTTTCAGAGAGCTGCT 1662
 501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
 1663 GATCTAAGATATGGCGCAATTCAGAGAGTGGAACTGCAATTTGCCCACTTGAAGGAAT 1722
 521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
 1723 TCTTCTGAAGAGATGTGATGCTCACAGAAACCTTGGCGCTGAAACATTTGCTGAGGTT 1782
 541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
 1783 GTGAGCCCTTGACAGGAGATTCAGTACGAGACTTGGCCAAATGAGAGGAGAGTTG 1842
 561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
 1843 ATTGGTCTTGTGATAGTTCATTAAGCGGTTTGGGACAGATCAAGCGGTAAATGCA 1902
 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 1903 GTTCTGAGGCAATTTAAGTCAAGGCGAGGACTTGGTAGGCGACACACGCAACTGGA 1962
 601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
 1963 TCATTTCTTATCTTGTGACCAACTGGTGTGGCAAACTGAGCTCGCCAGGCTTCTTGTCT 2022
 621 GluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGlu 640
 2023 GAGCAGCTGTTGATGATGAAACCTTCTAGTTCGATTGATATGCGAATATATGAA 2082
 641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 660
 2083 CAACACTCTGCTCTCGCTCATTTGGGCGACCCAGGAGGTATGTTGTCACGAGAGGT 2142
 661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680
 2143 GGCAACAACTAAGTGGCTGTGAGGAGGCGACCTTATTGTGTCATACTCTTTGATGAAGTG 2202

681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
 2203 GAGAAGGCTCATGTTGCTGCTTCAACACTCTGCTCCCAAGTTTGGATGATGCTGATTG 2262
 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSerAsn 720
 2263 ACAGACGGGCAAGGAGGACAGTTCGATTTCAGGAACCTCGGTGATANTCATGACATCAAC 2322
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2323 CTTGCTGTGAACACTCTCTTCAGGGCTTAAGTGGAAAGTAACAATGGAAGTGGCCCGG 2382
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
 2383 GACTGTGTGATCGGGAGGTGAGGAAACACTTCAGACAGAGCTCTTGAACAGGCTTCAC 2442
 761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
 2443 GAGATTGTGCTGCTCGACCCCTTTTCATGATGACCTTGGAGAAAGTAGCTCGGCTTCAA 2502
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2503 ATGAAAGAGCTTGTCTCGGCTTCTGTAAGAGAGGAGTTGCTTTGGCAGTCACTGATGCT 2562
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2563 GCTTTGGACTATATCTTGGCAGAGATTATGACCCGGTGTATGGTGTAGGCTTATAAGG 2622
 821 ArgTyrMetGluLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
 2623 AGATGATGAGAAAGGTGTTGACAGACTGTCTCAAGATGTTGTCGTCGAGAAATC 2682
 841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
 2683 GATGAAACTCCACTGTTTACATAGATGACGCGCTGCTGATCTTGTGTACCGGTAGAA 2742
 861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 2743 AGTGGAGGCTTAGTGGACGCTTCAACAGCAAGAGTCAGATGCTGCTGATTCATATTGCT 2802
 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMetArgIleGluGluIle 900
 2803 AACGGGCCAAAGAGAGTGTATGATGCTCAGGCGGTGAGAGATGAGGATCGAGGAATA 2862
 901 GluAspAspAspAsnGluGluMetIleGluAsp 911
 2863 GAAGATGACGATAATGAGGAAATGATCGAGGAT 2895

RESULT 2
 AX505540
 LOCUS
 DEFINITION Sequence 235 from Patent WO0216655.
 ACCESSION AX505540
 VERSION AX505540.1 GI:23386777
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 REFERENCE
 AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
 TITLE Stress-regulated genes of plants, transgenic plants containing
 same, and methods of use
 JOURNAL Patent: WO 0216655-A 235 28-FEB-2002;
 The Scripps Research Institute (US); Syngenta Participations AG
 (CH)
 FEATURES
 Source Location/Qualifiers
 1..2736
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 784 a 481 c 785 g 686 t

GIN

gment Scores:

d. No.: 1.05e-196 Length: 2736
re: 4545.00 Matches: 910
cent Similarity: 99.89% Conservative: 0
t Local Similarity: 99.89% Mismatches: 1
ry Match: 99.89% Indels: 0
Gaps: 0

09-812-350-17 (1-911) x AX505540 (1-2736)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGATCCAGAGAAATTCACACACAGACAAACAGACAAATTCCTACAGCTCATGAGCTA 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
61 GCTGTGAATGCAGGACATGCTCAATTCCTCTTGCATTTAGCTGGTGGTTCATCTCT 120
41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTCCGCTGGCGAGAACCGAGCTCAA 180
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
181 TCTGCTGAAGAGTGATCAATCAAGCTTTGAAGAAGCTTCCTTCACAATCTCCTCACT 240
81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaGln 100
241 GATGATATTCACGAGGATCTAGTCTTATTAAGGTCATTCGTCGCTCAAGCTCCTCAG 300
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
301 AAGTACAGAGGTGATACTCATTTGGCTGTGTGACCATGCTGATTTATGGCTCTTCTGAAGAT 360
121 SerGlnIleArgAspLeuLeuAsnGluValAlaGlyValAlaThrAlaArgValLysSerGlu 140
361 TCTCAATCAGGGAATTTGTTGAACGAAGTCGCTGTAGCGACGCGAGGGTAAAGCTGAG 420
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
421 GTTGAGAAGCTTCGTGGGAAGAAGGAGAAAGTTGAGAGTGTCTTCAGGGACACAAAT 480
161 PheGlnAlaLeuLysThrTrpGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
481 TTTCAAGCTTTAAAGACTATAGGAAGAGATTTGGTTGACAGACAGGAGGCTGATCTCT 540
181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
541 GTGATTTGGTCTGATCAGAGAGATTAGAAGAGTCTGAGGATTCCTTTCGAGGAGAACGAAG 600
201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
601 ACAATCCTGTCTTATTTGAGAGACCGAGGATTTGGTTAAACACAGCTGTGGTTGAGGTTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGGATTCGAAGAGAGATGTGCCCAACAGCTTACTGATGTGAGATTAATTTCCG 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGluGluArgLeu 260
721 TTGACATGGTGGCTTGTAGTTGTGGTGTCTTAAATACCGAGAGAGATTTGAAGAAAGGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
781 AAATCTCTTTTGAAGAAGTTTGAGAGCGCTGAAGGCAAAAGTGAATTCCTTTATGATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTTGGTCTTGTGTCTGGCAAACTGAAGGGTTCGATGATGATGATGATGATGATG 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320

901 AAGCCCATGTAGTAGGCGCAGCTTCGATGCAATTTGGTCTACACGCTTGAGGAATAC 960
321 ArgLysThrValGluLysAspAlaAlaPheGluArgPheGlnGlnValTrpValAla 340
961 AGGAATAATGTTGAGAAGATGCTGCTTTGAGAGAGGTTCCACAAAGCTCTATGTTGG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysThrGluGly 360
1021 GAGCCAAGTGTGCTGACACCAATTAGTATCTTTAGAGGACTCAAGGAGAATATGAGGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGTTGCGAATCCAAAGACAGAGCTTTATAAATGCTGCTCAGCTGTCTGCTCGT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATACTGCTGGCATTTTACCGGATTAAGCAATTTGATTTGCTGATGAGGCTTGTGG 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
1201 AATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTTGATTAACCTTGAAGGAAGAG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
1261 ATGCAAGTGGAAATTTGAATTCACGCTTGGAAAGGAGAGGATTAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGTGGGAAAGAGCTTGATGACCTGAGAGACAAGCTTCAGCCTCTCACG 1380
461 MetLysThrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
1381 ATGAAATACACAAAGGAGAAAGAGAGAAATTTGATGAGATTTCAAGGGCTTAAACAGAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgTrpAspLeuAlaArgAlaAla 500
1441 GAAGAGTCTATGTTTTCTTTGAGAGGAGGAGAGAGATGATGATGATGATGATGATGAT 1500
501 AspLeuArgTrpGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1501 GATCTAAGATATGCGCAATTCAGAAAGTGAATTCGCAATTCGCAATTCGCAATTCGCA 1560
521 SerSerGluGlnAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTCAAGAGATGTGATGCTCACAGAAACGTTGGGCTGAAACATTCGCTGAGGTT 1620
541 ValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
1621 GTGAGCCCTTGACAGAGGATTCAGTACAGAGCTTGGCCAAATGAGAGAGAGAGGTTG 1680
561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
1681 ATTGCTTGTGCTGATAGTTGCTAAGCGGCTTGTGGGACAGAAATCAAGCGGTAAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyValArgAlaGlnGlnProThrGly 600
1741 GTTCTGAGGCAATTCCTAAGGTCAAGGCGAGGACTTGGAAAGGCCAACAGCCCACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATTCCTTCTTGGACCAACTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1860
621 GluGlnLeuPheAspGluAsnLeuValArgIleAspMetSerGluTyrMetGlu 640
1861 GAGCAGCTGTTGATGATGATAAACCCTCTTAGTTGCGATTTGATGCTGCGAATATATGGA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTrpValGlyHisGluGluGly 660
1921 CAACACTGCTCTCTGCTCATTTGGGGCCACCCAGGATGATGTTGTTGCTCAGAGAGGT 1980
661 GlyGlnLeuThrGluAlaValArgArgProTrpCysValIleLeuPheAspGluVal 680

1981 GGACAACTAAGCTGAGGCTGTGAGGAGCGACCTATTGTGTCTATCTCTTTGATGAAGTG 2040

681 GlulysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeu 700

2041 GAGAGGCTCATGTGTCTGCTTCAACACTGCTCTCAAGTTTGGATGATGTCGATG 2100

701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleleMetThrSerAsn 720

2101 ACACAGGGCAGGAGGAGGAGTGTGATTTTCAGGAACCTCGGTGATATCATGACATCAAC 2160

721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740

2161 CTTGGTGTCAACACCTCTTCAGGGCTAACTGGGAAAGTAACATGGAAGTGGCCCG 2220

741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760

2221 GACTGTGTGATGCGGAGGAGGAGGAGTGTGATTTTCAGGAACCTCGGTGATATCATGAC 2280

761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780

2281 GAGATTGTGTGTGTCGACCCCTTTCATGACAGCTGTGAGAAAGTAGCTCGCTTCAA 2340

781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800

2341 ATGAAGACGTTGCTGTCGCTGCTGAAAGAGAGGAGTGTGTTTGGCAGTCACTGATGCT 2400

801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820

2401 GCTTTGGACTATATCTTGGCAGAGTGTATGACCCGCTGTATGCTGTAGCCCTATTAAGG 2460

821 ArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840

2461 AGATGGATGGAGAGAGAGTGTGACAGAACTGTCAAAGATGTTGTGCGTGAGGAAATC 2520

841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860

2521 GATGAAACTCCACTCTTATATAGATGACGCGCTGCTGTATCTTGTGACCGGGTAGAA 2580

861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880

2581 AGTGAGGCTCTAGTGAGCGCTTCAACAGCAGAGAGTCAAGATGCTGATTCATATTGCT 2640

881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900

2641 AACGGCCAAAGAGAGAGTGTGACAGCTCAGCGGTGAAAGAGATGAGGATCGAGGAAATA 2700

901 GluAspAspAsnGluGluMetIleGluAsp 911

2701 GAAGATGACGATAATGAGGAAATGATCGAGGAT 2733

ULT 3
89851
US
INITION
SSION
SION
WORDS
RCE
RGNISM
ERENCE
UTHORS
TITLE
JOURNAL
TURES
source

AX589851 2736 bp DNA linear PAT 24-JAN-2003
Sequence 33 from Patent WO02081695.
AX589851
AX589851.1 GI:27901086
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
Zhu, T., Glazov, E. A., Meins, F., Wang, X. and Chang, H. S.
Genes that are modulated by posttranscriptional gene silencing
Patent: WO 02081695-A 33 17-OCT-2002;
Syngenta Participations AG (CH)
Location/Qualifiers
1. 2736
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
784 a 481 c 785 g 686 t

ORIGIN

Alignment Scores:

Pred. No.: 1-05e-196 Length: 2736
Score: 4545.00 Matches: 910
Percent Similarity: 99.89% Conservative: 0
Best Local Similarity: 99.89% Mismatches: 1
Query Match: 99.89% Indels: 0
DB: 6 Gaps: 0

US-09-812-350-17 (1-911) x AX589851 (1-2736)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20

Db 1 ATCAATCCAGAGAAATTCACACACAGACAAACGAGACAATTCGTACAGCTCATGAGTA 60

QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40

Db 61 GCTGTGAATGCGAGGACATGCTCAATTCATCTCTTTGCAATTAGCTGGTGTGATCTCT 120

QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60

Db 121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTCCGCTGGCGAGAACGACGCTCAA 180

QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80

Db 181 TCTGCTGAAAGAGTGTATCAATCAAGCCTTGAAGAAGCTTCCTTCATCTCCTCCACT 240

QY 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100

Db 241 GATGATTTCCAGCGAGTTCTAGTCTTATTAGGTCAATTCGTCGTCTCAAGCTGCTCAG 300

QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120

Db 301 AAGTCACGAGGTGATATCTATTTGGCTGTGACCAAGTTCATTTATGTTGCTTCTTCAAGAT 360

QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140

Db 361 TCTCAATCAGGGATTTGTGAAACGAAGTCGCTGTAGCGACGCGAGGGGTAAAGTCTGAG 420

QY 141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160

Db 421 GTTGAGAGCTTCGTGGGAAAGAGGAAAGAAAGTTGAGAGTGCTTCAGGGGACACAAT 480

QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180

Db 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTTGATCT 540

QY 181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200

Db 541 GTGATTGGTGTGATGAGGAGATTAGAAGAGTGTGAGGATTTCTTCGAGGAGAGCAAG 600

QY 201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220

Db 601 AACATCTCTGTCTTATTTGGAGAGCCAGAGTTGTTGTAACAGCTGTGTTGAAGGTTTA 660

QY 221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgIleLeuSer 240

Db 661 GCACAAGGATTTGAAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCG 720

QY 241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyValPheGluGluArgLeu 260

Db 721 TTGGACATGGGTGCTGTAGTGTGCTGTGCTTAATACCGAGGAGAGTTTGAAGAAGGTTG 780

QY 261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280

Db 781 AATCTGTTTGAAGAGTTGAGAGCTGTGAGGCAAGTGAATCTCTTTATTGATGAG 840

QY 281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaLeuPhe 300

Db 841 ATTCATTTGCTTCTTGTGCTGCGCAAACTGAAGGGTGTGATGATGATGATGATGATGAT 900

QY 301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluTyr 320

901 AAGCCATGTTAGTACAGGCGACGCTTCGATGCTTACCAAGCTTGAAGATAC 960
 321 ArgLysTyrValGluLysAspAlaPheGluArgPheGlnValTyrValAla 340
 961 AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGCTATGTTGCG 1020
 341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
 1021 GAGCCAAAGTGGCTGACCACTTAGTATCTTATAGAGACTCAAGAGAGATATGAGGGA 1080
 361 HisHISGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeuSerAlaArg 380
 1081 CATCATGCTGTGGCAATCAAGACAGAGCTCTTATAATGCTGCTAGCTGCTGCTCGT 1140
 381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
 1141 TACATAACTGCTGGCAATTTACCGGATAAAGCAATGATTTGGTGTGATGAGGCTTGGCG 1200
 401 AsnValArgValGlnLeuAspSerGlnProGluIleAspAsnLeuGluAlaGlyArg 420
 1201 AATGTGAGAGTCCAGCTTGATGCTCACTCAAGAGATGATTAACCTTGAAGAGAGAGG 1260
 421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerIleAla 440
 1261 ATGCAGCTGGAAATTTGAATTCAGCCCTTGAAAGGAGAGAGGATAAAGCCAGCAAGCT 1320
 441 ArgLeuIleGluValArgLysGluLeuAspLysLeuArgAspLysLeuGlnProLeuThr 460
 1321 CGACTTATAGAGGTGCGAAGAGCTTGATGACCTGAGAGAGATGATTAACCTTGAAGAGAGG 1380
 461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgGluLysGlnLysArg 480
 1381 ATGAATACAGAGAGAGAGAGAGAGAGATTCATGATTCGAGGCTTAAACAGAAAGA 1440
 481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
 1441 GAAGAGCTCATGTTTCTTTCAGGAGGAGAGAGAGATGATGATTCGAGGCTTAAACAGAAAGA 1500
 501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
 1501 GATCTAAGATATGGCGCAATCAAGAGTGGAACTCTGCAATTTGCCCACTTGAAGAGAACT 1560
 521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
 1561 TCTTCTGAAGAGATGCTGATCTCACAGAAAGCTTGGCCCTGACACATCTGCTGAGGTT 1620
 541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
 1621 GTGAGCCGTTGACAGGGATTCAGTGCAGAGACTTGGCCCAAAATGAGAGAGAGAGGTTG 1680
 561 IleGlyLeuAlaAspArgLeuHisLysValValGlyGlnAsnGlnAlaValAsnAla 580
 1681 ATGTGCTTGTGCTGATAGGTTGATTAAGCGGGTTGTTGGACAGATCAACAGCGTAAATGCA 1740
 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 1741 GTTCTGAGGCAATTTAAGTCAAGGCGAGGACTTGAAGGCCCAACAGCCAACTGGA 1800
 601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
 1801 TCATCTTATTCCTTGGACCAACTGCTGTTGCAAAACTGAGCTCGCCCAAGGCTCTTGTCT 1860
 621 GluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGlu 640
 1861 GAGCAGCTGTTTGTATGATGAAACCTCTTAGTTCGATTCGATATATATATATGGA 1920
 641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 660
 1921 CAACACTGCTGCTCTGCTCATTTGGGGCACCACAGGAGTATGTTGGTTCAGCAGGAGGT 1980
 661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680

1981 GGACAACAACTAAGTGGCTGTGAGGCGGACCTTATTTGTTGTCATACCTCTTTGATGAAGTG 2040
 681 GluLysAlaHisValAlaValPheAsnThrIleuLeuGlnValLeuAspAspGlyValGlu 700
 2041 GAGAAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCCAGGTTTTGGATGATGCTCGATTG 2100
 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
 2101 ACAGAGCGGCAAGGCGAGCAGCTCGATTTCAGGAACCTCGGTGATATCATGACATCAAC 2160
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2161 CTTGGTGTGAAACACCTCTTTCAGGCGCTAACTGGGAAGTAACAATGGAAGTGGCCCGG 2220
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
 2221 GACTGTGATGCTGGAGGAGTGAGGAACACTTCAGACCGAGCTCTTGNACAGGCTTGAC 2280
 761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
 2281 GAGATTGTGCTGTGACCCCTTTCACATGACCACTGAGGAAAGTAGTCTCGGCTTCAA 2340
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2341 ATGAAAGACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2401 GCTTTGACTATATCTTGGCAGAGATTATGACCGGCTGATGCTGCTGCTGCTGCTGCT 2460
 821 ArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
 2461 AGATGATGAGAGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
 841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
 2521 GATGAAACTCTCCACTGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
 861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 2581 AGTGGAGGCTTAGTGGACGCTTCAACAGGCAAGAGTCAAGATGCTGCTGCTGCTGCTGCT 2640
 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
 2641 AACGGGCCAAAGAGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
 901 GluAspAspAspAsnGluMetIleGluAsp 911
 2701 GAAGATGACGATATGAGGAATGATCGAGAT 2733

RESULT 4
 AF218796
 LOCUS
 DEFINITION
 Arabidopsis thaliana 6376 bp DNA linear PLN 13-APR-2000
 cds.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 6376)
 Hong S.W. and Vierling, E.
 Mutants of Arabidopsis thaliana defective in the acquisition of
 tolerance to high temperature stress
 Proc. Natl. Acad. Sci. U.S.A. 97 (8), 4392-4397 (2000)
 20226114
 PUBMED
 10760305
 2 (bases 1 to 6376)
 Hong, S.-W. and Vierling, E.
 Direct Submission

Db	3591	ACATCAAA	CCTTGGTGTGAA	CACCTCTCTTGCAGGGCTAA	CTGGGAAGATACAA	TGGAA	3650									
Qy	738	Valala	ArgAspCys	ValMet	ArgGlu	ValArg	GlyHis	Phe	Arg	Pro	Glu	Leu	Leu	Asn	757	
Db	3651	GTGGCCGG	GACTGTGTGATCGG	AGGTGAGGAACA	CACCTTCAGAC	CAGAGCTCTTGAAC	3710									
Qy	758	ArgLeu	AspGlu	IleVal	ValPhe	AspPro	LeuSer	His	Asp	Gln	Leu	Arg	Gly	Val	Ala	777
Db	3711	AGGCTT	GACAGAGT	TGGGTTCG	ACCCCTTTCAT	GACAGTCAG	TCAGTTCGGAAG	TAGCT	3770							
Qy	778	ArgLeu	GlnMet	His	AspVal	AlaVal	ArgLeu	AlaGlu	ArgGly	ValAla	LeuAla	Val				797
Db	3771	CGGCTT	CAAAATGA	AGACGT	TGCTGT	CGCGCTTCT	GAAAGAG	GAGT	TGCTTTGG	CAGTC	3830					
Qy	798	Thr	AspAla	alaLeu	aspTyr	IleLeu	AlaGlu	SerTyr	AspPro	Val	-----	813				
Db	3831	ACTGAT	GCTGCTTTGG	ACTATATCTTTGG	CAGAGATTATG	ACCCGGT-AA	GCTCTTCAC	3889								
Qy	813	-----	-----	-----	-----	-----	-----	813								
Db	3890	ATGATG	CAGAAATGC	ATATCAT	TGTTGATCTCT	CAGCGTGGTTGT	TGCTGTGAG	CTGACT	3949							
Qy	814	-----	-----	-----	-----	-----	-----	-----	822							
Db	3950	TTGTGG	TGCTTCTTGT	TTTGTGTGGT	TAGTGTAT	GTGCTAGGCGCTA	TAAGGAGATGG	4009								
Qy	823	Met	Glu	Leu	Leu	ValThr	GluLeu	SerLeu	SerVal	ValArg	GluGlu	Ile	Asp	Glu	842	
Db	4010	ATGAGA	GAAGAAGGTGGT	GACAGAACTGT	CAAAAGATGGT	TGTGCGTGA	GAGAAATCGATGA	4069								
Qy	843	Asn	SerThr	ValTyr	IleAsp	AlaGly	AlaGly	AspLeu	ValTyr	ArgVal	GluSer	Gly	862			
Db	4070	AACTCC	ACTCTTTACAT	ATAGATGC	AGGCGCTGGT	GTATCTTGTGTAC	CGGCTAGAAAGTGGGA	4129								
Qy	863	Gly	Leu	Val	AspAla	SerThr	GlyLeu	SerAsp	Valleu	IleHis	IleAla	AsnGly	882			
Db	4130	GGTCT	ATAGTAC	CGCTTCAAC	AGCAAGATCG	ATGATGCT	GTGATTCATAT	TGCTAAC	CGGG	4189						
Qy	883	Pro	lys	ArgSer	AspAla	IleGln	AlaVal	LeuSyls	Met	ArgIle	GluGlu	IleGlu	Asp	902		
Db	4190	CCAA	GAGAGAGT	GTATGC	AGCTCAGG	CGGTGAAG	AGATGAGGATCG	GAGGAATAGA	GAT	4249						
Qy	903	Asp	Asp	AsnGlu	GluMet	IleGlu	Asp	911								
Db	4250	GAGC	ATAATCAG	GAATATGATC	GAGGAT	4276										

RESULT 5	AC020579	50821 bp	DNA	linear	PLN 19-JAN-2001
LOCUS	AC020579				
DEFINITION	Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence, complete sequence.				
ACCESSION	AC020579				
VERSION	AC020579.5	GI:12324896			
KEYWORDS	HTG.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1. (bases 1 to 50821)				
AUTHORS	Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.				
TITLE	Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence				
JOURNAL	Unpublished				
REFERENCE	2. (bases 1 to 50821)				
AUTHORS	Lin, X. and Kaul, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-JAN-2000) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA, xlin@igr.org				
REFERENCE	3. (bases 1 to 50821)				
AUTHORS	Town, C.D. and Kaul, S.				
TITLE	Direct Submission				


```

HYHEVPRSGHFPHYTKGVVDDIVKSLTSDVRAASRVC1"
gene      23489..24554
/contig="F1017.6"
/notes="similar to ribosomal protein L33B GB:NP_014877 from
[Saccharomyces cerevisiae]"
mRNA      join(<23489..23520,23750..23785,24176..24404,24499..24554)
/contig="F1017.6"
CDS       join(23489..23520,23750..23785,24176..24404,24499..24540)
/contig="F1017.6"
/codon_start=1
/product="putative ribosomal protein; 23489-24540"
/protein_id="AAG52401.1"
/db_xref="GI:12324899"
translation="MVKGRQGRVRLVYRGITLIGYKRKSNQYPTSLVQIEGVNTQF
EVNWKGRMAVYIYAKTKNGSHVRCIWKVTRPHGNSGVVRAKFTSNLPKSNKSR
VRVFWYPSNI"
gene      25509..29950
/contig="F1017.7"
/notes="similar to GB:K1AA0361 from [Homo sapiens]"
mRNA      join(<25509..25911,25995..26563,26663..29047,
29144..29950)
/contig="F1017.7"
CDS       join(25509..25911,25995..26563,26663..29047,29144..29950)
/codon_start=1
/product="putative phosphoribosylformylglycinamide
synthase; 25509-29950"
/protein_id="AAG52403.1"
/db_xref="GI:12324901"

Alignment Scores:
Seq. No.:      2,78e-188      Length:      50821
              4388.00      Matches:      908
              86.56%      Conservative: 0
              86.56%      Mismatches: 3
              96.44%      Indels:      140
              8          Gaps:      4

09-812-350-17 (1-911) x AC020579 (1-50821)
1 MetAspProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
13093 ATGAATCCAGAGAAATTCACACAGACAGACAAATTCACAGCTCATGAGCTA 13152
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
13153 GCTGTGAATGCAGGACATGCTCAATTCCTCTTGCATTAAGTGGTGTGATCTCT 13212
41 AspProThrGlyLeuPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
13213 GATCCACCGGTATATTCTCAAGCAATCTCTAGTCCCGGTGGCGAGAACGACGCTCAA 13272
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
13273 TCTGCTGAAGAGTGAATCAATCAGCCTTGAGAGCTTCCTTCAATCTCTCCACT 13332
81 AspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
13333 GATGATATCCAGCGAGTCTAGTCTTATTAAAGTCAATTCGTCGTCCTCAAGCTGCTCAG 13392
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp 120
13393 AAGTCAGAGGTGATACTCATTTGGTGTGTGACCAAGTGTATGATGAGTCTCTTGAAGAT 13452
121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
13453 TCTCAATCAGCGATTTGTTGAACGAGTCTGTTAGCGACGCGCGAGGGTAAGTCTGAG 13512
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
13513 GTTGAGAAGCTTCGTGGGAAAGAGGAAAGATTGAGAGTGTCTTCAGGGGACACAAAT 13572
161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180

```

```

Db      13573 TTTCAAGCTTTAAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAAAGCTGTATCT 13632
Qy      181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200
Db      13633 GTGATTCGTGTGATGAGGAGATTAGAAAGAGTCTGAGGAGTCTTTCGAGGAGAACGAAG 13692
Qy      201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
Db      13693 AACAACTCTGTCTTATTGAGAGCCAGGAGTTGGTAAACACAGCTGTGGTTGAAGGTTTA 13752
Qy      221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
Db      13753 GCACAAAGAGATTGAAAGAGAGATGCCCCAACAGCTTACTGATGTGAGATTAATTCG 13812
Qy      241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu 260
Db      13813 TTGACATGGGTGGTGTAGTGTCTGTGGCTAAATACCGAGAGAGTTTGAAGAAAGTTG 13872
Qy      261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
Db      13873 AAATCTGTTTGAAGAGATTGAGGAGCTGAAGCAAGTGAATCTCTTTATTGATGAG 13932
Qy      281 IleHisLeuValLeuGlyAlaGlyLysThrGlySerMetAspAlaAlaAsnLeuPhe 300
Db      13933 ATTCAATTGGTTCCTTGGTGTGCTGCAAACTGAAGGGTCGATGATGATGATCTGTTTC 13992
Qy      301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
Db      13993 AAGCCCAATGTTAGCTAGAGGCGAGCTTCGATGATGATGATGATGATGATGATGATGATG 14052
Qy      321 ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrValAla 340
Db      14053 AGGAAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTCCAAACAAAGTCTATGTTGG 14112
Qy      341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
Db      14113 GAGCCAAAGTGTGCTGACACCATTAGTATCTTCTTAGAGGACTCAAGGAGAGTATGAGGGA 14172
Qy      361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
Db      14173 CATCAAGTGTGGAATCCAAAGACAGAGCTTTTAAATAGCTGCTCAGCTGTGCTCGT 14232
Qy      381 TyrIleThr----- 383
Db      14233 TACATAACTGATGTTAAGATTCTTAATCCTAAGCTGATGTTTATGTTTTCATATAGTG 14292
Qy      384 -----GlyArgHisLeuProAspLysAlaIle 392
Db      14293 TTGTTCTTGACATGACATTTTGTGGGATTTAGTGGGATTTACCGGATTAAGCAATT 14352
Qy      393 AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu 412
Db      14353 GATTTGGTTCATGAGGCTTCTGCGAATGTGAGAGTCCAGCTTGTAGTCAACCTGAAGAG 14412
Qy      413 IleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArg 432
Db      14413 ATTGATAACCTTGAAGAGAGAGATGACGTCGAGAAATTTGAACCTTCACGCTTGGAAAG 14472
Qy      433 GluLysAspLysAlaSerLysAlaArgLeuIleGluVal----- 445
Db      14473 GAGAAGGATAAGCCAAAGCTCGACTTATAGAGGT-ATGATACAGCTTTCTTTCTTC 14531
Qy      445 ----- 445
Db      14532 ATCAGTAGATAGCGCTTCAAAATATAAGACTATTGCTGTTCTCTGGAATTTCTTGATGTT 14591
Qy      445 ----- 445
Db      14592 AGCATCTGTTTTCGATGAGTCTTTGATATAATTTTCTCAAACTATGATCTGTAGGTG 14651
Qy      446 ArgLysGlnLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLys 465

```

Qy	798	ThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProVal	----- 813
Db	15792	ACTGATGCTGCTTTGGACTATATCTTGGCAGAGAGTTATGACCCGGT-AAAGTCTCTTCAC	15805
Qy	813	-----	813
Db	15851	ATGATGCAGAAATGCATATCATTTGTTGATCTCTCAGCGTGGTTTGTTCGTGACAGTCTC	15910
Qy	814	-----TyrGlyAlaArgProIleArgArgTrrp	822
Db	15911	TTGTGGTCTCTTTCTTTGTTTCTGTTAGGTGTATGTTGCTAGGCCCTATAAGGAGATGG	15970
Qy	823	MetGluLysGlyValValThrGluLeuSerLysMetValValArgGluGluIleAspGlu	842
Db	15971	ATGCGAAGAAGCGTGGTGACAGACTGTCAAAGATGGTTGTGCGGTGAGAGAAATCGATGAA	16030
Qy	843	AsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSerGly	862
Db	15031	AACTCCACTGTTTACATAGATGCAGCGCTGTTGATCTGTGTACCGGTTAGAAGTGA	16090
Qy	863	GlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAlaAsnGly	882
Db	16091	GGTGTAGTGGACGCTTCAACAGCGCAAGAAGTCAGATGTGCTGATTTCATATTGCTAACCGG	16150
Qy	883	ProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIleGluAsp	902
Db	16151	CAAAAGAGAAGTGTATGTCAGCTCAGCGCGTGAAGAAGATGAGATCGAGGAATAAGAAGAT	16210
Qy	903	AspAspAsnGluGluMetIleGluAsp	911
Db	16211	GAGCATATGAGAAATGATCGAGAT	16237
RESULT 6			
SOYSB100HS			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
FEATURES			
source			
gene			
CDS			

LFIDEHLVLGAGRTGEGMDANLFPKMLARGQLRCIGATTLEBYRYKVEKDAAFERR
FOOVFAEPSVDVTSILRLGKERVEGHGVRIODIRLMAAOLSNRYITGRHLPDKA
IDLVDVACANVRVQDSPEIDNLERKEMQLEVEHLHLEKEKDAKARLVEYRGAL
DLDRKLOQVLMKRYKEREVEDEIRLKKREELLFALQEABRRIDILARAADLYKAL
CEVETAIQOLSGSTEENLMETITVPEQIAEVVSRWTGIPVTRIGONEKERLIGLGR
LHSRVGQDQAVNAVEARSLRAGLRPOQPTGSLFGPTGVKTELEAKALAEOLF
DNEOLVRIDMSYMECHSVSRLLGAPPGYVHBEQGLTEAVRRPVSIVLFPDEVK
AHTSVNTLLQVLDGRITDQGRITVDFRNTVIIWISNLGAHLLSLGSGKCTWQVAR
DRWQEVROPRPELLNBLDELIVDPDLSDHOLKVAELOMKDVASRLAEKGLAVT
DHALYILSESVDYFGARPIRWLEKKVTELSMLVREIDENSTVYIDAGNGGE
LVYRKNKGWVNPPTGQKSDILIQIPNGHAPKTADVAOVKKMKLEIDDDDEMESE

BE COUNT 850 a 564 c 863 g 772 t

GIN

gment Scores:
d. No.: 3049
Length: 3049
Matches: 4003.00
Conservative: 72
Similarity: 94.85%
at Local Similarity: 86.97%
Mismatches: 39
Indels: 8
Gaps: 6

09-812-350-17 (1-911) x SOVS100HS (1-3049)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
136 ATGAATCCTGAGAAGTTTACTCACAAGACTAATGAAGCTTGTGCGCGGCCACGAGCTC 195
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
196 GCGATGAGTTCAGGCGCACCGCAATTGATCCCATCCACTTGGCCCGCGCTGATTTCC 255
41 AspProThrGlyIlePheProGlnAlaIleSerSerAla--GlyGlyGluAsnAlaAla 59
256 GATCCCAACGGCATCTTCGCTAGCGATAACAGCGCGCGCGCGGAGGAAATCGGCA 315
60 GlnSerAlaGluArgValIleLeuAlaLeuLysLeuProSerGlnSerProPro 79
316 CGCGCGTGGAGCGAGTGTGAACAGGCTCTCAAGAGCTACCTGCCAGTCCCTCCG 375
80 ProAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAla 99
376 CGGACGAGTGGCGGAGCAGCACCACTCTGTGAGGCGCATCAGGAGGACAGCGCGG 435
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuGlu 119
436 CAAATATCAGTGGCGACACGCGTTTGGCGGTGATCAGTTCATCTCGAATCTCGAA 495
120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
496 GACTCCCAATCGAGACCTTGTGAAGAGCGGGGTTGCGGTGGCGAGGATGAATCG 555
140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
556 GAATGATAGCTTCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 615
160 AsnPheGlnAlaLeuLysThrTyrrGlyArgAspLeuValGluGlnAlaGlyLysLeuAsp 179
616 AATTTCCAGCTTTGAAGACTTATGCGCGTGACCTTGTGTAACAGCGGGAAGCTCGAC 675
180 ProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThr 199
676 CCTGTTATTGGCGGTGACGAGAGATTTAGAAGGGTTGTGAGGATTTCTATCAGCGAGG 735
200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
736 AAGACACACCGGTTCTCGTTGGAGAACCGGGGTGTGGGAAACCTCGGTTGTGGAGGG 795
220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
796 TTGGCAGACAGGATAGTAAAGAGGGATGTTCCCAAGCAACCTTCTGATGTGAGGCTATT 855
240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrrArgGlyGluPheGluArg 259

856 GCCTTGGATATGGGGCGTGTGTCGCGGTGCCAATATAGAGCGGAGTTTGAGGAGCGG 915
260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp 279
916 TTAAGAGCTGTTTGAAGGAAGTGGAGGAGCGCTGAGGGAAGGTGATATCTTTCATTGAT 975
280 GluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu 299
976 GAGATTTCATTGCTGCTTGGTGTGTGAGAACTTGAAGGCTCCATGCTGCTGCTATCTA 1035
300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlu 319
1036 TTCAAACCTATGCTTGTGCGCGCCAGCTTAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1095
320 TyrArgLysTyrrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrrVal 339
1096 TACAGGAAGTATGTGGAGAAGGATGCTGCATTCGAGGAGGAGTTCCACAGGTTTTTGIG 1155
340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLysTyrrGlu 359
1156 GCGGAACCTAGTGTGTTGATACCATTCATTCATTCCTGCTGCTGCTGCTGCTGCTGCT 1215
360 GlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAla 379
1216 GCGCATCACGCTGTAGAAATTCAGGACCGCTGCTTGTGTTATGCGAGCTCAATTTGCTAAC 1275
380 ArgTyrrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
1276 CGGTATATACCTGGCGCTCATCTCTGACAGGCAATTCATTTGTTGACGAGGCTTGT 1335
400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
1336 GCAATGTAGGCTTCAACTTGTATGATGCTGAGCTGAGGAAATTCATAACCTTGAAGAGAG 1395
420 ArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysValAspLys 439
1396 AGAATGAGCTAGAGTGAAGTGAAGTTCATGCTCTGAGAGAGAGAGAGAGAGAGAGAG 1455
440 AlaArgLeuIleGluValArgLysGluLeuAspLysAspLeuArgAspLysLeuProLeu 459
1456 GCGCGCTTGTGAGAGTGGAGAAAGTTCATGATGCTTGAAGGACAGCTTTCAGCTTGT 1515
460 ThrMetLysTyrrArgLysGluLysGluArgIleAspGluIleArgArgLysGlnLys 479
1516 ATGATGAAGTACCGAAAG 1575
480 ArgGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrrAspLeuAlaArgAla 499
1576 AGAAGAGAGCTTCTTTTGTCTTACAGAGGCTGAGAGAGATATGATCTGGCTAGAGCT 1635
500 AlaAspLeuArgTyrrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
1636 GCAGACCTGGATATGAGCAATTCAGAGGTGAGAACTGCAATACCAACCTTGAAGGG 1695
520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
1696 ---AGCACTGAAGAGATCTGATGTTGACTGAAACTGTTGGACCGGAGCAATAGCTGAG 1752
540 ValValSerArgTyrrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg 559
1753 GTTGTGAGCGCTGGACCGGTATACCGATTACAGGCTTGGCCAAATGAAAGAGAGAGAG 1812
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
1813 TTGATTGAGCTTGTGTGACAGATTGACAGCAGCAGAGTTGTAGGACAGCAAGCTGTTAAT 1872
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
1873 GCTGTGCTGAGGCTGTGTTGAGATCTAGAGCTGGGTAGAGAGAGCTCAACCAACCACT 1932
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619

221 GATCAAAATGCTATTTTCGACAAAGGATGTCATGCTGGTGGGAATGAAGAAGTAGCT 280
60 GlnSerAlaGluArgValIleAenGlnAlaLeuLysLysLeuProSerGlnSerPro 79
281 AATTCAGTGGAGCGGCTATTGAATCAAGCGATGAAGAAGTACCTTCTCAACACCGGCT 340
80 ProAspAspIleProAlaSerSerSerLeuLysValIleArgAlaGlnAla 99
341 CCGAGCAATCCCACTAGCACTTCCCTTATCAAGGTGTTACGCCGAGCAATCGTCG 400
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
401 CAGAAGTCTCGTGGTGACAGCCATTTAGCAGTGGATCACTGATTTTAGGACTGCTAGAA 460
120 AspSerGlnIleArgAspLeuAenGluValGlyValAlaAlaThrAlaArgValLysSer 139
461 GACTCCCAATTCGAGATCTTTTGAAGAAGCTGGAGTGGATGAGTGCATCAAGAGTGAATCA 520
140 GluValGluLysLeuArgGlyLysGluLysValGluSerAlaSerGlyAspThr 159
521 GAGGTAGAGAACTTATAGAGAAAGGAAGGAAGTGAAGTGAAGTCTTCAGGGACACC 580
160 AsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLysLeuAsp 179
581 ACATTCAGACACTCAACACTTATGCGCGTGTCTTGTGGAACAAGAGTGGAT 640
180 ProValIleGlyArgAspGluLysLeuArgValValArgValAlaIleLeuSerArgThr 199
641 CCGGTGATGTTAGGAGTGAAGAATTTAGAGAGTCTGTTCCGATTTTATCAAGGAGGACT 700
200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
701 AAGAACACCCCGGTTCTTATGGAGAGCCGCTGCTGGTAAACACAGCTTGTGAAGGG 760
220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
761 CTAGCACAGAGGATGTACGTGTGTATGTCCTCAAGTAAATTTAGCTGATGCTTATA 820
240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGluPheGluGluArg 259
821 GCATTTGGATATGGAGCGCTAGTGTCTGGAGCTTAAGTACAGAGGTGAATTTGAAGAGG 880
260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp 279
881 CTGAGGCTGTGCTGAAGAAGCTTGAAGAAGCGAGGAAAGTAACTATCTTTCATTGAC 940
280 GluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu 299
941 GAGATACATTTAGTCTCTCGTCTGCTGGTGGAGAGGCTTATGATGCTGCTAATCTG 1000
300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlu 319
1001 TTTAAGCCAAATGCTAGCAGAGGTCATTAACGGTGCATTTGGTCAACTACACTCAGAGGAG 1060
320 TyrArgLysTyrrValGluLysAspAlaAlaPheGluArgPheGlnGlnValTyrrVal 339
1061 TACAGGAAGTATGTTGAGAGGATGCTGCTATTTGAGAGGCGTTTCCAGCAGGTGTATGTT 1120
340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyrrGlu 359
1121 GCTGAGCCTAGTGTGCTGACACATTTAGTATTTCTCCGTGGTGTGAAGGAGAGATGAA 1180
360 GlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAenAlaGlnLeuSerAla 379
1181 GGGCATCATGTTGTCAAAATTTAGAGACAGAGCTTTGTAGTGGCTGCCAGCTCTCATCT 1240
380 ArgTyrrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
1241 CGGTACATTACAGTCTGACATCTGCCAGATGAAGCTATTGACCTTAGTTGATGAAGCTTGT 1300
400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
1301 GCAATGTTAGAGTTTCACTTGTATGATCACTGAGGAAATTTGAATCTTTGAGAGGAG 1360
420 ArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLys 439
1361 AGAATTCAGCTAGAGTTGAACTTTCAGCTCTCAGAGAGGAAAGACAAAGCTAGCAAA 1420
440 AlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeu 459
1421 GCACGTCTCTAGAGTGAAGAAAGAACTTGTATTTGAGAGACAACTCCAACTTTG 1480
460 ThrMetLysTyrrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLys 479
1481 ATGATGAGGTACAAAGAAAGAGAGGATAGATGAGCTGCCAGGCTCAAGCAAAAG 1540
480 ArgGluLeuLeuMetPheSerLeuGlnGluAlaGluArgTyrrAspLeuAlaArgAla 499
1541 CGCATGAGCTCATCTATGCTTTACAAGAGCTGAAAGGAGATATGATCTGGCAGGGCA 1600
500 AlaAspLeuArgTyrrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
1601 GCAGATCTGAGATATGGGCAATTTCAAGAAAGTGAAGAACTGCAATAGCAATCTTGAGAGT 1660
520 ThrSerSerGluGluAenValMetLeuThrGluAenValGlyProGluHisIleAlaGlu 539
1661 ACCTCAGCT---GAAAGTACATGCTTAACAGAGACTGTGGTCTCTGATCAGATTGGGAA 1717
540 ValValSerArgTyrrThrGlyIleProValIleThrArgLeuGlyGlnAenGluLysGluArg 559
1718 GTTGTGAGTCTGTGGACTGTGTTTCCGCTCTCAAGGCTTGTGTCAGATGAGAAAGAA 1777
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAenGlnAlaValAsn 579
1778 CTGATTTGCTTGGCGCATAGATTGCACCAAGAGTGTGGGCAAGATCATGCGATTAGA 1837
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
1838 GCTGTGCTGAAGCCGTGTGAAGTCCAGAGCTGTGTTAGGAGGCCACAGCAACCACT 1897
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
1898 GGTTCATTCCTTTCTTGGGCGCAACTGTGTTGGAAAGACAGAGCTCGCTAAAGCTCTT 1957
620 AlaGluGlnLeuPheAspAspGluAenLeuValArgIleAspMetSerGluTyrrMet 639
1958 GCAGAGCAGCTCTTTGATGATGATAAATGATGATCAGATAGACATGTCGAGTACATG 2017
640 GluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrrValGlyHisGluGlu 659
2018 GAACAACTCTGTTTCCGCTGATTTGCTCCACAGGTTATGTTGGCATGATGAG 2077
660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrrCysValIleLeuPheAspGlu 679
2078 GGAGGACAACTTACTGAAGCTGTTAGGAGCGGCTTACAGTGTGTTGCTATTATTTGAGAA 2137
680 ValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArg 699
2138 GTTGAAGAAGCCATCTGCGAGTGTTHATACATGCTTCAAGTCTTGATGAGGAGG 2197
700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAenSerValIleIleMetThrSer 719
2198 TTAAACAGATGGTCAAGCGCGCACAGTGTGATTTCCCAACTCCGTGATTTATGACTTCA 2257
720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
2258 AACTTGGGACAGAGTATCTGTTGTCTGATTTAATGGGCAATGTACCATGGAGCAGCT 2317
740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu 759
2318 CCGTAATGCTCATGAGAGGTGCAAGACAGATTTAAGCCCGAGCTGCTGAAATCGGCTG 2377
760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
2378 GATGAGATTGTTGTTGATCTCTGTCGCCACAGAGTGTGAGGCAAGTATGCGCTAC 2437

```

780 GlnMetIysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
|||||
2438 CAGATGAAGACGCTTGCACCTACGCTGGCTGAGAGGGTATTGCATTGGCGCTTACTGAG 2497
|||||
800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProfile 819
|||||
2498 GCAGCTCCAGATGCTACTACTACAGAGAGATTATGACCGGTTTATGGTGAAGACCTATT 2557
|||||
820 ArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
|||||
2558 AGAGATGGTTGGAGAGAGAGTGGTACCGAGCTATCCAGATGCTCGTGAAGGAGGAG 2617
|||||
840 IleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuValTyr 857
|||||
2618 ATTGATGAGAACTCAACCGTTTACATAGATGCTGGGTACGCGGGAAGATCTAAACCTAC 2677
|||||
858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeu 876
|||||
2678 AGGTGGAGAGAGAGAGGCTTGTGAATGCTGCCACCGGCAAAATCTGATATATTG 2737
|||||
877 IleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArg 896
|||||
2738 ATTCACTCTCTAATGTGCTCC--AGAGTGTGCTGTCCAGCAGTCAAGAGATGAGG 2794
|||||
897 IleGluGluIleGluAspAspAspAsnGluGlu 907
|||||
2795 ATTGAAGAAATTGAAGATGACCAATGGAAGAT 2827
|||||

JLT 8
AF133840 3084 bp mRNA linear PLN 21-NOV-2001
US Zea mays heat shock protein HSP101 (HSP101) mRNA, complete cds.
INITIATION
AF133840
MISSION
AF133840.1 GI:4928487
WORDS
Zea mays
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 3084)
Young, I.E., Ling, J.J., Geisler-Lee, C.J., Tanguay, R.L., Caldwell, C.
and Gallie, D.R.
Developmental and thermal regulation of the maize heat shock
protein, HSP101
Plant Physiol. 127 (3), 777-791 (2001)
21563054
11706162
2 (bases 1 to 3084)
Young, T.E. and Gallie, D.R.
HSP101 Diversity in Monocot Species
Unpublished
3 (bases 1 to 3084)
Young, T.E. and Gallie, D.R.
Direct Submission
Submitted (09-MAR-1999) Biochemistry, University of California,
Watkins Drive, Riverside, CA 92521, USA
Location/Qualifiers
1..3084
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
1..3084
/gene="HSP101"
145..2883
/gene="HSP101"
/feature="HSP101"
/codon_start=1
/product="heat shock protein HSP101"
/protein_id="AAD33606.1"
/db_xref="GI:4928488"
/translation="MNPNFTHKNEATVGAHEITAVEAGHAQLTPLHLAVALADKGG
ILRQAITGASGDGAGAGSFERVLNLSLKLPSPDPDSVFASSTALIKVIRRAQSAQ

```

```

KRGDSHLAVDOLLGLLEDSQISDLCKEAGVSAARVAEAELEKRGGERVESASGD
TNFQALTYGRDLVEQACKLDPVIGROBEIRVRVILSRRTKNNPVLIGEPGVGTAV
VEGLAQIRVGRDVPNSLIDVRLIALDMGALVAGAKYGEFEERLKAIVLEKEBAEKG
ILFDEIHLVAGAGTEGSDAANLFKPMALGAGQRCIGATTLEBYRYKVEKDAFER
RFQGVFAEPSVQDTSVTLRGKEKEGHGVRIQDRALVVAALSAARYIMGRHLPOK
AIDVDEACANVRVQDSQPEIDNLEKRIQLEVELHALKEKDKASKARLIVRKE
LDDLKILQPLTMKTRKEKRIEIRKLQKREBLQFLQEAERMDLARVADLYKGA
LOEIDAAISKLESETGENMLTETVGPQIAEVSVRWTGIPVIRLQGNKRLVGLAD
RLHQVQVQTEAVASAVAEAVLRAGLRPOOPTGSGFLGPTGVGKTELAKALAEQL
PDENLLVRIWMSEMEQSARVLIAGPPVGHGEGQLTEQVRRRYSVILFDEVE
KARVAFNTLLQVLDGRLTDQGRITVFRNTVIMTNSLGAELHLAGMVGKNSKVA
RDLVMQVRRHRPELNLDELIVDFPLSEOLRKLVARLQMKDVAVLAERGIATLAV
TDAALDLSLSYDPVYCARPIREKRVYVTLQSEMLIQEIDENCTVYIDAAPKXD
ELIVYKDRSGSLVNAETGMSDILLIQVNSSTRSDAAQAVKMKRIMEDEDEGMDEE"
BASE COUNT 670 a 874 c 1057 g 483 t
ORIGIN

Alignment Scores: 8.25e-167 Length: 3084
Pred. No.: 3883.50 Matches: 772
Score: 92.88% Conservative: 76
Best Local Similarity: 84.56% Mismatches: 58
Query Match: 85.35% Indels: 7
Gaps: 6
DB: 8

US-09-812-350-17 (1-911) x AF133840 (1-3084)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 145 ATGAATCCGGACAACTTCAACCCACAAGACGAAGAGCGATCGTGGGGCGGCAGAGAT 204
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
DB 205 GCGGTGGAGCGCGCGCAGCGAGCTCAACCGCGCTGCACCTGCGCGCGCGAGTGTGCTGCG 264
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
DB 265 GACAGGGCGCGCATCTCGCGCAGCGCATCAACGGGGGGTTCGGGGGGCGCAGCGAGCGCC 324
QY 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
DB 325 GGGGACTCTGTCGAGCGCGTGTGAACAACTCGCTCAAGAGCTCGCTCGCGTGGCGCG 384
QY 79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAla 98
DB 385 CCGCGCGACTCTCGTCCCGCGCTCCAGCGCAGTCAAGAGTCAATCGCGCGCGCGAGTCC 444
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
DB 445 GCGCAGAAAAACGCGGGGACTCTCACTCGCGCTGCACCGAGTGTGCTCGCGCTGCTC 504
QY 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
DB 505 GAGGACTCGCAGATCTCCGACTCGCTCAAGAGGCGCGCGTGTGCGCGCGCGCGGTCGCG 564
QY 139 SerGluValGluLysLeuArgGlyLysGlyLysLysValGluSerAlaSerGlyAsp 158
DB 565 GCGGAGCTTGAGAGCTCCGCGCGCGGGAGGCGCGCGCTGAGTTCGCGTGGCGGAGAC 624
QY 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
DB 625 ACCAATTCAGCGCGCTCAAGACATACGCGCGGGACCTGTTGAGCAGCGCGGAGCTG 684
QY 179 AspProValIleGlyArgAspGluLysLeuArgValValArgIleLeuSerArgArg 198
DB 685 GACCCCGTTCATCGCGCGCAGCAGGAGATCCCGCGCTGTGCGCATCTCTCTCGCGCGCG 744
QY 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
DB 745 ACTAAGAACAAACCCCTCTCTCAATCGCGCGAGCGCGCGCTGTGGCAGACGCGCGCTG 804
QY 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238

```

599	QY	ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyValThrGluLeuAlaLysAla	618
1942	Db	ACTGGCTCGTTCCTCTGCTCGGTCGACATGGCGTGGGGAACCTAGCTAGCCAAAGGCC	2001
619	QY	LeuAlaGluGlnLeuPheAspAspGluAenLeuLeuValArgIleAspMetSerGluTyr	638
2002	Db	CTAGCCGAACACAGCTGTTGCAGCAGCAGAACCTTCTTGTCGCGCATCGACATGTCGGAGTAC	2061
639	QY	MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGlu	658
2062	Db	ATGGAGCAGCACTCGGTTGCTCGCTCATCGGAGCACCCGCTGGCTACGTGCGCCCATGA	2121
659	QY	GluGlyGlyGlnLeuThrGluAlaValArgArgArgProTyrCysValIleLeuPheAsp	678
2122	Db	GAGGGTGGGCAGCTGACTGAACAGAGTGAAGAGGAGCCGTACACGCTGATCTCTGTTTCGAC	2181
679	QY	GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly	698
2182	Db	GAGTTCGAGAGGCCCATGTCGCCGTGTTCAACACCCTGCTCCAGGTCTCTCGACGACGGC	2241
699	QY	ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThr	718
2242	Db	AGGTTGACGGATGGGCAAGCGCAGCAGCAGCTTTCAGGAACAACCGCTGATCATCATGACA	2301
719	QY	SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal	738
2302	Db	TCAACCTCGGCCCGCCGAGCACCTTCCTCGTGGGATGGTGGGCAAGAACTCCATGAAGGTC	2361
739	QY	AlaArgAspCysValMetArgGluValArgLysPheArgProGluLeuLeuAsnArg	758
2362	Db	GCTCGCGCATCTGGTCATGACGAGAGGTGAGAGGCACTTCGACCCGAGCTGCTGAACCGT	2421
759	QY	LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg	778
2422	Db	CTCGACGAGATCGTGATCTTCGATCTCTGTCGCCACGAGCAGCTGAGGAAGGTCGCTCGC	2481
779	QY	LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr	798
2482	Db	CTTCAGATGAAGGATGTGGCCGTCCGCTTTCGGAAGGGGCAATCGCTCTGGCTGTGACC	2541
799	QY	AspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro	818
2542	Db	GACGCCGCAATTGACATCATCTTGCTCTCTTACGATCCGTTGATACGGCGCGCGCCA	2601
819	QY	IleArgArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu	838
2602	Db	ATCAGGAGTGGATCGAGAGAGGGTGGTCAACAGCTGTGCAAGATGCTGATCCAGGAG	2661
839	QY	GluIleAspGluAenSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuVal	856
2662	Db	GAGATCACGAGAACTCACCGGTCTCATCGACGCCGCCGCCGCAAGGACGAGCTGGTC	2721
857	QY	TyrArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal	875
2722	Db	TACAGGTGACCGGACGCGCGTCTGGTGAACGCTGAGACGGGGATGAAGTCGACATC	2781
876	QY	LeuIleHisIleAlaAenGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMet	895
2782	Db	CTGATCAGGTCCCCAACAGCTCCACCAGGAGCGACCTCGCAGGCGCGTCAAGAAGATG	2841
896	QY	ArgIle---GluGluIleGluAspAspAsnGlu	907
2842	Db	AGGATCATGGAGGAGACGAGGACGGTATGACGAGGAG	2880

RESULT 9
AF332981
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

AF332981 3148 bp mRNA linear PLN 02-JAN-2002
Oryza sativa heat shock protein (HSP101) mRNA, complete cds.

AF332981
AF332981.1 GI:18033449

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 3148)
Agarwal, M., Sahi, C., Young, T., Agarwal, S., Mitra, S., Ganesan, K.,
Agarwal-Katayar, S., Gallie, D.R. and Grover, A.

Isolation and Analysis of HSP101 from Rice

Unpublished

2 (bases 1 to 3148)

Young, T.E. and Gallie, D.R.

Direct Submission

ITIE

Submitted (02-JAN-2001) Biochemistry, University of California,

Boyce Hall, Riverside, CA 92521-0129, USA

Location/Qualifiers

1..3148

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

1..3148

/gene="HSP101"

135..2873

/gene="HSP101"

/codon_start=1

/product="heat shock protein"

/protein_id="AA157165.1"

/db_xref="GI:18033450"

/translation="MNPNDFHFKNEALVAHEIASBAGHAQLTPLHLVAALADKGG
ILRLAISQAGDAGAPDSFVRSVVAALKLPSPPDPSPASTALIKVIRRAQSAQ
KRGDSHLAVQLLGLLEDSLI SDCLKEAGVSAARVAELKELRGEGEKRVSASGD
TNFQALKTGRDLVFEQAGKLPVIGRDEIRRVVRIISRTKNNPVLIGPGVGKTA
VLEDEHLVLAGAGTSGESDAANLEKPMARGOLRCITGATTLVEYKRYVEKDAEPK
RFQVFAEVPDITISILRGLKEKYECHGVRIQDRALVVAQSLSARYIMGHELPK
AIDLVEACANRVQDSPEIDNLEKRIQLVEHHALEKEDKASKARLVEKKE
LDDKDLQPLTKMYRKEDEIRKLKORREBELQFTLOEARMOLARVADLUKGA
LQEDIVAKLESETGNMLETVTGPEQIAEVVSRMTGIPVTRIGQNDKERLVGLAD
RLHORVQAQASAVAEVARSAGLRPOQPTGSLFLGPTGVGKTELAKALAEOL
FDENLLVIRDMSEVMECHSVAELIGAPGVYGHGEGQLTEQVRRPYPVILPDEVE
KAHVAFNTLLQVLDLGGTGTGQGTVDPRNTVIMTSLNLAHLLAGVGNKSMKVA
RDLVQEVRRHFRHRLNRLDEIVDFLSHEQLRKVRLQMKDVAVRLAERGVALLV
TDAALDVLSLSYDPVYGAVIRWIKRVVTLQSKMLIOEIDENCTVYIDAAHPKD
ELAVRVDNRGGVLNAETQKSDIILTOVFNAGATGSDAAQAVKQRIWEDDEGMDER"

644 a 931 c 1087 g 486 t

E COUNT

GIN

gment Scores:

d. No.:

re:

cent Similarity:

t Local Similarity:

ry Match:

7.34e-165

3840.50

92.00%

83.79%

84.41%

8

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

3148

765

75

66

7

5

09-812-350-17 (1-911) x AF332981 (1-3148)

1 MetAnProGluLysPheThrHisLysThrAnGluThrIleAlaThrAlaHisGluLeu 20

135 ATGAATCCGAGCAACTTCACGACCAAGACCAACAGAGGCGCTCGTGGCGGCGCACGAGATA 194

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40

195 GCGTCGAGCGGGCACGCGAGCTGACCGCTGACCTGCTGGCGGCGTGGCGGCG 254

41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59

255 GACAAAGCGCGGATCTCCGAGCGGCGATCTCGACGCGTCCGCGGCGACCGCGCGCG 314

60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78

315 CCGGACTCGTTCGAGCGCGTGGTGGAGCGGCGGCGCTCAAGAAAGTGCCTCGCAGTCGCGG 374

79 ProProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
375 CCGCGCGGACAGCGTGCCTCGCGCTCCAGCGCGCTGATCAAGGTTCATCGCGCGCGAGTTC 434
99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
435 GCGCAGAAAGCGCGGAGCTCCCACTTCGCGCGTCCAGCCAGTCTGCTGGGCTCTCTG 494
119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
495 GAGGACTCTCTAATCTCCGACTCCCTCAAGAGGCGCGCTGCTCCGCGCGCGGTGGCG 554
139 SerGluValGluLysLeuArgGlyLysGluLysLysValGluSerAlaSerGlyAsp 158
555 GCGAGCTCGAGAAGCTCGTGGCGGAGGCGCGCAAGGTTCGCGCTCCGCGCGCGAC 614
159 ThrAspPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluAlaGlyLysLeu 178
615 ACCAATCTCCAGCGCTCAAGACCTACGCGCGCGACCTCTCGAGCAGCGCGGTAAAGTTC 674
179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
675 GACCCGCTCATTTGGCGCGACGAGGAGATCCGCGCGTCTGCGCATCTCTGCGCGCGC 734
199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGlu 218
735 ACAGAAACAACCCCGTACTCATCGCGAGCGCGCGTCCGCAAGACGCGCTCTGTTGGAG 794
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
795 GGCCTCGCGAGCGCATCTGCGCGGAGACGTGCCAGCAACCTGCTGACCTCGCGCTC 854
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 258
855 ATCGCGCTCGACATGGCGCGCTCTGCGCGCGCGCAAGTACCGCGCGCGAGTTCGAAGAA 914
259 ArgLeuLysSerValLysGluValGluAlaGluAspAlaGlyLysValIleLeuPheIle 278
915 CGGCTCAAGCGCGTCTCAAGAGGTGGAGGCGCGCGAGGTGATCTCTTTCATC 974
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
975 GAGCAGATACCTCTGCTTCGCGCGCGGAGGACGAGGCGGTCCATGACGCGCGCAC 1034
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
1035 CTGTTCAAGCAATGCTCGCGAGGCGAGTCTCGGTGTCATCGCGCGCACGCGCTGGAG 1094
319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgPheGlnGlnValTyr 338
1095 GAGTACCGCAAGTACGTCGAGAAGGACGCGCGTTCGAGAGGCGGTTCAGACAGGTGTT 1154
339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysLysGlyTyr 358
1155 GTCCGAGGCGGAGCGTCTCCGAGCACCATCAGCATCTCTCGAGGCTCTCAAGGAGAGTAC 1214
359 GluLysHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
1215 GAGGCGCACCATCGTGTGAGGATCCAGGACCGCGCTCTGCTCGCGCGCGAGCTCTCG 1274
379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1275 GCGAGGTACATCATGGGACCGCACCTGCTGTCGCAAGAGCATGACCTCTGTCGAGAGGCC 1334
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluIleAspAsnLeuGluArg 418
1335 TCGCGCAACGTGAGGTGTCAGTGGACGACGACGAGGAGATCGACACCTTCGAGAGG 1394
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1395 AAGAGGATTCAGCTGGAGGTGGAGCACCGCTGCGTGGAGAAAGGAGGAGGAGGAGGAGC 1454
439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458

2532 GAGCCGCGGTGGACGCTATCTTCCTCCTATGATCGGTTCATGCCGCGGAGGCGG 2591

819 ILeArgTrrpMetGlulysLysValValThrGlulSerLysMetValValArgGlu 838
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
2592 ATCAGGAGGTGGATCGAGAAGAGGGTGCTGACGCAGCTGTCCAAGATGCTCATCCAGGAG 2651

839 GluIleAspGluAsnSerThrValTyrlleApeAlaGly-----AlaGlyAspLeuVal 856
||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
2652 GAGATCGACGAGAATCGCACGGGTGTACATCCAGCGCGCGCCGCAACAGACGAGCTGGCC 2711

857 TyzArgValGluSer---GlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
2712 TACAGGGTGGACAACC CGCGGGGGTTGGTGAACGCGGACGCGGCGCAGAAAGTCGACATC 2771

876 LeuIleHisIleAlaAsnGlyProLysArg---SerAspAlaAlaGlnAlaValLysLys 894
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
2772 CTCATCCAGGTGCCAACCGCGCGCCGCCACCGGAGCGACGCCGCGAGCGGTGAAGAAG 2831

895 MetArgIleGluIleGluAspAspAenGluGlu 907
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2832 ATGAGGATCATGGAGACGACGAGCGGCATGGACGAGAG 2870

RESULT 10
AF097363
LOCUS AF097363
DEFINITION Triticum aestivum heat shock protein 101 (Hsp101b) mRNA, complete cds.
ACCESSION AF097363
VERSION AF097363.1 GI:4558483
KEYWORDS Triticum aestivum (bread wheat)
SOURCE Triticum aestivum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 3107)
AUTHORS Campbell,J.L., Kluvea,N.Y., Zheng,H.G., Nieto-Sotelo,J., Ho,T.D. and Nguyen,H.T.
TITLE Cloning of new members of heat shock protein HSP101 gene family in wheat (Triticum aestivum (L.) Moench) inducible by heat, dehydration, and ABA(1)
JOURNAL Biochim. Biophys. Acta 1517 (2), 270-277 (2001)
MEDLINE 21240210
PUBMED 11342108
REFERENCE 2 (bases 1 to 3107)
AUTHORS Campbell,J.L., Kluvea,N.Y. and Nguyen,H.T.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Plant and Soil Science, Texas Tech University, Mail Stop 2122, Lubbock, TX 79409, USA
LOCATION/Qualifiers Location/Qualifiers
FEATURES
source
1..3107
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Mustang"
/db_xref="taxon:4565"
/dev_stage="10 day old seedling"
/note="treated 2 hrs. at 37 degrees C."
gene
1..3107
/gene="Hsp101b"
CDS
112..2868
/gene="Hsp101b"
/note="ClpB family member"
/codon_start=1
/product="heat shock protein 101"
/protein_id="AAD2629.1"
/db_xref="GI:4558484"
/translation="MNPDNFTKTNELVAHEASAEAGHQITPLLAALAADKSGQILRQAVAGSGNAGSDSFERVISGLAKKLPSQPPDPSVPASTALLIKAPRAQSQA
XKRGDHLAVDLQLMGLLDDAQIDLCLKEAGSVSRARELEKLRGGNSNRKVESAG
DTNFALKTVGRDLVEVAGLKDPVIGRDSEIRRVRIISRKNPVLIIGPGVGKTAFL
VEGLARVVRGVDPNLVDRLVALDMGALVAGYRGFEERKLVIVKEVEDAEAF
VLIDIEHLIVLAGRTSGMSDAANLFKMLARGQLRCIGATTLEEVKYVVEKDPE
EPOOVYVASEPVANTSILEGTEVKGHGCVIRIDRAIVVAOALSARYIMGHSLPD

KAIDLVDACANVRVOLDSPBEIDNLERKRIQLEVEHLEKEKDKASKARLIVEVRK
 ELDLRLKQLPMTWKYRKEKRIEIRKLKORRELOFTQDEARRMDLARVADLVKG
 ALQEDLAIALEGTEGNMLMTETGPEQIAEVVSWTGTPTVRLQNDKERLVGMA
 DRLHTRVVGQTEANAVAEVIRAGLRFPQPTGSLFUGPTGSGVGTTELAKALABQ
 LFDHNLVIRIDMSEYMEHQHVAELIIGAPPYVGHHEGGQLTEQVRRRPSYVILFDEY
 EKARHVAFTLIDLDLDRITDGOORTVDFRNTVIIMTSLNGAEHLHLAGVMGNKSV
 ARDLVMOQVRHFRPELILNDEIVFDPPLSHEOLRVARLQMDVAVRLAERGVALA
 VTDAALDLSILNADVPVGPASPIRWLEKRVVLTOLSKWMLIOEIDENCTVYIDAANKD
 ELAYRDRSGGLVUNNETQORSIDILLQVFNALGSGKGGEAAKAVKMKRWEDGDDEGM
 DEDV"

605 a 954 c 1116 g 432 t

gment Scores:

id. No.: 3107
 Length: 3107
 Matches: 760
 Conservative: 82
 Mismatches: 66
 Indels: 12
 Gaps: 7

09-812-350-17 (1-911) x AF097363 (1-3107)

1 MetAspProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 112 ATGACCCCGGACAACTTCACGCACACAGACCAAGAGCGCTGCTGCGCGGCGCACGAGCG 171
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisIleAlaGlyAlaLeuIleSer 40
 172 GCGTCGAGCGCGGCACGCGCAGATCACGCGCTGCACCTGCGCGCGCGCTGCGCGCG 231
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
 232 GACAACTCGGGCATCTCGGCACAGCGCGCTGCGGGGCGCTCGCGGGGACGCGTCCGCG 291
 60 ---GlnSerAlaGluArgValIleAsnAlaLeuIleLysLeuProSerGlnSerPro 78
 292 GCGGACTCGTTCGAGCGCTGCTCTCCGCGCGCTCAAGAACTGCGCGTCCGCGTCCGCC 351
 79 ProProAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAla 98
 352 CGCGCGGACTCGTTCGCGCTCCACGCGCTCATCAAGGCCATCGCGCGCGCGAGTCC 411
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 412 GCGCAGAAAGCGCGGACTCCCACTTCGCGCTCGCACAGCTGCTCATGGSCCTCTCTC 471
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 472 GAGGACGCGCAGATCGCGCTCCCTCAAGAGGCGCGGCTGTCGCGCTCGCGGGTGC 531
 139 SerGluValGlnLysLeuArgGlyLysGlu---GlyLysValGluSerAlaSerGly 157
 532 GCGGACTCGAGAAGCTCGCGCGCGGGAACAATCGCGCAAGTCTGAGTCCGCGCTCCGCG 591
 158 AspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLys 177
 592 GACACCAACTTCAGCGGCTCAAGACGACGCGCGGACTCTGTCGAGGTGGCGCGCAAG 651
 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
 652 CTCGACCCGCTCATCGCGCGCAGAGGAGATCGCGCGCTGTCGCGCATCTCTCTCACGC 711
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 712 CGCACCAAGAACACCCCGCTCTCATCGCGAGCCCGCGGCTGGGCAAGCGCGCTCGTG 771
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 772 GAGGGCTTCGCGAGCGCGCTGTCGCGCGGGAAGTGCACCAAGCACTGCTCGACGTCGCG 831
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGluPheGlu 257

832 CTGTCGCGCTCGACATGGCGCGCTCTGTCGCGCGGCCCAAGTACCGCGCGGAGTTCGAG 891
 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 892 GAGCGCTCAAGCCCTGCTCAAGGAGGTGGAGGAGCGCGAGGCAAGTTCATCTCTTC 951
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 952 ATCGACGAGATACCTCTGCTCTGCGCGCGGAGGAGCGAGGGTCTCATGGACGCGCC 1011
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317
 1012 AACCTGTTCAGCCGATGCTTCGCCAGGGGCGAGCTCCGGTGCATCGCGCCACACGCTG 1071
 318 GluGluTyrglyTyrglyValGluLysAspAlaAlaPheGluArgArgPheGlnVal 337
 1072 GAGGATACCGCAAGTACGTCGAGAGGAGCGCGCTTCGAGCGCGGTTCACAGAGGTG 1131
 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 1132 TAGTGGCCGAGCGCGCTCGCGCACCATCAGCATCTCTGCGAGGCTCAAGGAGAG 1191
 358 TyrGluGlyHisGlyValAlaGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 1192 TACGAGGGGACCAACGCGCTGCGCATCCAGGACCGCGCATCTGTCGCGCGCGAGCTC 1251
 378 SerAlaArgTyrglyThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 1252 TCGCGAGGTACATCATGCGCGCGCACCTTCGCGACAGGCCATCGACCTGGTGGACGAG 1311
 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluIleLeuAspAsnLeuGlu 417
 1312 GCGTCGCCCAACGTGAGGTGCGAGTGGACAGCGCGCGAGGAGATCGCAACCTGGAG 1371
 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 1372 AGGAAGCGGATTCAGCTGGAGGTGAGCTGCACGCGCTGGAGAGGAGAGGAGCAAGGCC 1431
 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 1432 AGCAAGCGCGCTGTTGGAGGTGAGGAGGTGAGGAGGTGAGGAGGTGAGGAGGTGAG 1491
 458 ProLeuThrMetLysTyrglyArgLysGluGluIleAspGluIleArgArgLeuLys 477
 1492 CCGCTGACGATGAAGTACAGAGAGGAGAGCGGATCGACGAGATCCGGAAGCTGAAG 1551
 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrglyAspLeuAla 497
 1552 CAGCGCGCGAGGAGCTGCGAGTTCACGCTGCGAGAGCGGAGCGCGCGCATGGACCTG 1611
 498 ArgAlaAlaAspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 1612 CGCGTCGCGACCTCAAGTACGCGCGCTCCAGGAGATCGACGCGCGCATCGCCAGACTG 1671
 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 1672 GAGCGCGAGACCGCG---GAGACCTGATGCTCCAGACCGCTCGCGCGCGCGAGCATC 1728
 538 AlaGluValValSerArgTyrglyThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 1729 GCGGAGGTGGTGAACCGGTGACCGGATCCCGGTGACCGGCTCGCGCGAGAGAGCAAG 1788
 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 1789 GAGCGCTCTGCGCATCGCGGACAGCGCTGCACACGAGGCTGGTGGCGGCGAGCGAGGCC 1848
 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 1849 GTGAACGCGCTCGCGAGGCGGTCTCGAGGTCCAGGCGCGGCTTGGCGCGCGCGAGCAG 1908
 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
 1909 CCCACCGGCTCTCTCTCTCTGCGCGCGCGCGCGCGGTGGGCAAGACCGAGCTCGCCAG 1968

618 AlaLeuAlaGluGlnLeuPheAspGluAsnLeuLeuValAlaGleAspMetSerGlu 637
 1969 GCCCTCGCGGAGCAGCTGTTGACCAAGAACCTGCTGTCGCGCATCGCATGTCGAG 2028
 638 TyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProGlyTyrValGlyHis 657
 2029 TACATGGAGCAGCACTCGCTCGCGGCTCATCGGAGCGCCACCTGGCTATGTCGGGAC 2088
 658 GluGluGlyGlyGlnLeuThrGluAlaValAlaArgArgProTyrCysValLeuPhe 677
 2089 GAGGAGGCGGCGGACCTGACGAGCAGGTGAGAGGAGCGGTACAGCGTATCTGTTTC 2148
 678 AspGluValGluLeuAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 2149 GACGAGGTGGAGAGCGGACGTCGGGCTGTTCAACAGCTCTCCAGTGTCTGGAGCAG 2208
 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAlaPheArgAsnSerValLeuMet 717
 2209 GGCGGCTGACCGAGCGGCAAGCGGAGGAGCGGTGAGCTTCGGAACACCGGTGATCATG 2268
 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
 2269 ACTTCAACCTCGCGGCGGAGCAGCTCTCGCGGAGTGGTGGGAGCAACTCCATGAG 2328
 738 ValAlaArgAspCysValMetArgGluValAlaArgLysHisPheArgProGluLeuLeuAsn 757
 2329 GTCCGCGGAGCTCGTCTCATGAGGAGGTGAGGAGCACTTCAGGCGGAGGTGCTGAAC 2388
 758 ArgLeuAspGluLeuValPheAspProLeuSerHisAspGluLeuArgLysValAla 777
 2389 CGCTCGGAGCAGATCGTCTGTCGACCGGCTGTCGCGGAGTGGTGGGAGCACTGAGGAGTCTGCT 2448
 778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
 2449 CGCTCGGAGTGGAGGAGTGGGCTGCGGCTCGCGGAGCGGCGGCTCGCTCGGCGTCC 2508
 798 ThrAspAlaAlaLeuAspTyrLeuAlaGluSerTyrAspProValTyrGlyAlaArg 817
 2509 ACCGAGCTCGCTGAGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2568
 818 ProLeuArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArg 837
 2569 CGCATCAGAGGTGGATCAGAGCGGCTGTCGACGAGCTGTCGAGAGTGTGATTCAG 2628
 838 GluGluLeuAspGluAsnSerThrValTyrLeuAlaGlyAlaGlyAsp---LeuVal 856
 2629 GAGGAGTTCGAGAGCACTGACGCTGATCATCGACGCGCCCAACAGCAGCTGGCC 2688
 857 TyrArgValGlu---SerGlyGlyLeuValAlaPheAlaSerThrGlyLysLysSerVal 875
 2689 TACCGGCTGAGCGGAGCGGCGGCTGTCGACGCGGAGCGGCGGAGCGGCGGAGTCCGACATC 2748
 876 LeuLeuHisLeuAlaGlnGlyPro-----LysArgSerAspAlaAlaGlnAla 891
 2749 CTATCAGGTGCGGAGCGGCGGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 2808
 892 ValLysLysMetArgLeuGluGluLeuGluAspAspAsnGluLeuMetLeuGluAsp 911
 2809 GTGAGAGATGCGGCTC-----ATGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGAT 2862
 AFI74433 3058 bp mRNA linear PLN 14-MAY-2002
 Triticum aestivum heat shock protein 101 (HSP101c) mRNA, complete cds.
 AFI74433 GI:6013195
 Triticum aestivum (bread wheat)
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.
 1 (bases 1 to 3058)
 Campbell, J.L., Klueva, N.Y., Zheng, H.G., Nieto-Sotelo, J., Ho, T.D.
 and Nguyen, H.T.
 Cloning of new members of heat shock protein HSP101 gene family in
 wheat (Triticum aestivum (L.) Moench) inducible by heat,
 Biochim. Biophys. Acta 1517 (2), 270-277 (2001)
 21240210
 MEDLINE
 11342108
 PUBMED
 2 (bases 1 to 3058)
 Klueva, N.Y., Campbell, J.L., Zheng, H. and Nguyen, H.T.
 Direct Submission
 Submitted (02-AUG-1999) Plant and Soil Sciences, Texas Tech
 University, Mail Stop 2122, Lubbock, TX 79409, USA
 Location/Qualifiers
 1..3058
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultiivar="Mustang"
 /db_xref="taxon:4565"
 1..3058
 /gene="HSP101c"
 94..2835
 /gene="HSP101c"
 /note="CIPB family member"
 /codon_start=1
 /product="heat shock protein 101"
 /protein_id="AA01280.1"
 /db_xref="GI:6013195"
 /translation="MNPDKFTHTNEALAAAHMASEAGHAQITPLHAAALAAADRSG
 ILQATAHAGGNDAAERFVASALKLPSPDPPTTASTSLKAVRAQSAQ
 KSGDSHLAVDQMLGLEDPOISDALXGANGISAARVKAIEKLGDNRRVESAGD
 TNGQALTVRGDVLVAGKLDVPIGRDEIRRVIRLSRTKNNPVLIGEPGVGTAV
 VEGALVRVGRDVEGSLNLDVLDVMDGALVAGYRGEFEELKAVLKEVEBEAGV
 ILTDETHVLGAGRTGGMDAANPKMLRQGLRCIGATTLTEYRKYVEKDAFER
 RFOVFVAPSPDVTSLILGLKEKVEHGVRIQDRALVIAAQLSSRYIMRHLDPK
 AILVDEACANVQLDSQPEIDNLERKIOLEVEHLEKEKDKAKARLVDRKE
 LDDRLKPLQMKYRKERIDEIRSLKORRELQTLQEAERMDLARVADLVGA
 LQEDRAIAKLEGTENMLTETVGPDDQAEVVSRTGTIPVRLQNEKARLIGLAD
 RLQRVVVGQYEAIVAVGEAVIRAGLRPQPTGFLGPTGVGTGKELAKALASQL
 FDENLIVRIMSEYMEQSVARLIGAPPYVGBEGGQLTEQRRRYSVILDFVE
 KAHVAVNTLLVDLDRGLTDGRTVDFRNTVIIMTSLGAHILAGVMGSMKVAR
 DLYWQVRRFRPELLNRLDEIVIFDLSHEQLRKVARLQMKDVAVLAERGVALVT
 DAALDVILSLUSTDPYVGAFIRMIKRIKIVTELSKMLIREIDENSTVVIDAPSKDE
 LTVGVDRKHGGLVNAHTGKSDILIQVPSGAVGDAHAHVKKIMQDSVEVDMEEE"

BASE COUNT 680 a 836 c 1025 g 517 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.62e-164 Length: 3058
 Score: 3822.50 Matches: 761
 Percent Similarity: 91.47% Conservative: 75
 Best Local Similarity: 83.26% Mismatches: 69
 Query Match: 84.01% Indels: 9
 Ds: 8 Gaps: 6
 US-09-812-350-17 (1-911) x AFI74433 (1-3058)
 QY 1 MetaAnProGluLysPheThrHisLysThrAsnGluThrLeuAlaHisGluLeu 20
 Db 94 ATGATCCCGACAGTTCACGACACCAATGAGCGCTCGCGCGCGCTCGAGATG 153
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
 Db 154 GCGTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 213
 QY 41 AspProThrGlyLeuPheProGlnAlaLeuSerAlaGlyGlyGluAsn-----Ala 58
 Db 214 GACAGTGGCGGATCTCTCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 273
 QY 59 AlaGlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78


```

799 AspAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818
|||||
2488 GAGCGCGCCTGAGATCATCTGTCACCTGTTACGATCCGGTCTATGGCGCCAGGCCA 2547
|||||
819 IleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
|||||
2548 ATCCGAGATGATCGAGAGAGGATAGTACGGAGCTCTCCAGATGTTGATCCGCGAG 2607
|||||
839 GluIleAspGluSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuVal 856
|||||
2608 GAGATCGAGAGAACTCCACGGTGTACATCGACGCTGCGCCAGCAAGAGCAGCTGACC 2667
|||||
857 TyrArgValGluSer---GlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
|||||
2668 TATGGCTGCAAGACAGCGAGGGCTGGTGAACGGCGGACAGCGGCCACAGATCCGACATC 2727
|||||
876 LeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMet 895
|||||
2728 CTGATCCAGGTTCTTAGCGGAGCTGTTGGGGCGATGCGCGCAGCGCGCTGAAGAAGATG 2787
|||||
896 ArgIle-----GluGluIleGluAspAspAsnGluGlu 907
|||||
2788 AAGATCATCGAGACAGCGAGAGGTGGACGACATGGAGGAA 2829
|||||

RULT 12
883344
US
Triticum aestivum heat shock protein 101 kDa (HSP101) mRNA,
complete cds.
AF083344
AF083344.2 GI:11561807
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 2821)
Wells, D.R., Tangway, R.L., Le, H. and Gallie, D.R.
HSP101 functions as a specific translational regulatory protein
whose activity is regulated by nutrient status
Genes Dev. 12 (20), 3236-3251 (1998)
9784498
2 (bases 1 to 2821)
Gallie, D.R., Tangway, R.L. and Wells, D.R.
Direct Submission
Submitted (10-AUG-1998) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
3 (bases 1 to 2821)
Gallie, D.R., Tangway, R.L. and Wells, D.R.
Direct Submission
Submitted (06-DEC-2000) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
SEQUENCE
Sequence update by submitter
On Dec 6, 2000 this sequence version replaced gi:3982832.
LOCATIONS
Location/Qualifiers
1..2821
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
1..2821
/gene="HSP101"
65..2821
/gene="HSP101"
/codon_start=1
/product="101 kDa heat shock protein"
/protein_id="AAC83689.2"
/db_xref="GI:11561808"
translations="MNPDNFTHTKNEALVAHAASEAGHAQITPLHAAALAAADKSG
ILKQAVAGSGNASAGDSFERYVLACALRLPSQSPDPSVASTALIKAIRAQSAG
KFGDHLAYDQLMGLLEDAQADCLKEAGVSASVRAELDKRGDMSRKVSAG
DTTFQALTKYGRDLVEVAGKLPVIGRDEIRVRILSRRTKNNPVLIGFVGKTA

```

```

VVEGLAQRVVRGDPVPSNLLDRLVALDMGALVAGAKYRGFEERLKAIVLKEVEEADGK
VILFDIEHLVLGAGRTESMDAANLFKPMARGQLRCIGATTLEBYRYKVEKDAPE
RRFOQYVAPESVADPTISILGLKEVGHGVR IODRAIVVAOQSARVIMGRHLPD
KADLVEACRVRVOLDSPSEIDNLERKIQLEVELHALEKEDKASKARLVEVRK
ELDDRLDKLOPLTMKRBKRIIDEIRKLQRRELOFTLQEAERMDLARVADLKYG
ALQIEDAIAKLEGETGENMLTETVGPQIAEVSRSWTGIPVIRLGQNDKRLVGM
DRLHTRVVGTEAVNAVALRSLRAGLGPQPTGFLGPTGVGTETELAKALABQ
LPDENLLVRIDMSVMEQHSVARLIGAPPVYVHEEGQLTEQVRRRPSVILFDEV
EKAHAVFWTLLOVLDGRLTDGQRTVDPRNTVIIMTNSLGAHLLAVMGVNSMKV
ARDLVQEVRRHRPPELLNRLDENVIFDPLSHEOLEKVARLQKMDKOVLAERGVLA
VTDAALDVLISLAYDPVIGARPRIWRKRVITQLSKMLIQEIDENCTVIYDAADKD
ELAYRVDRSGLVNAETQSRSDILIQVPNGALGGGGGGAAKAVKKRVMEDGDEDM
DEDV"
BASE COUNT 524 a 902 c 1035 g 360 t
ORIGIN
Alignment Scores:
Pred. No.: 1,24e-162 Length: 2821
Score: 3790.00 Matches: 752
Percent Similarity: 91.09% Conservative: 86
Best Local Similarity: 81.74% Mismatches: 70
Query Match: 83.30% Indels: 12
DB: 8 Gaps: 7
US-09-812-350-17 (1-911) x AF083344 (1-2821)
QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 65 ATGAACCGGACAACTTCACGACACAGACCAACGAGCGCTGGTGGCGGCGCAGAGGCG 124
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
DB 125 GCGTCGAGGCGCGGCCACGCGCAGATCACCGCTTGACCTGCGCGCGCGCTGGCGGCG 184
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
DB 185 GACAAGTCCTGTCGAGCGCGCTCGCGGAGCGCTCGCGGGGCGTCCGCGGGAATGTCGCGG 244
QY 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
DB 245 GGGGACTCGTTCGAGCGCGTCTCGCGCGCGCGCTCAGGAAGCTCGCGTCCGAGTCGCGC 304
QY 79 ProProAspIleProAlaSerSerSerIleLysValIleArgArgAlaGlnAla 98
DB 305 CCGCGGACTCGTTCGCGCGCTCCACGCGCTCATCAAGGCCATCCGCGCGCGCGCGAGTCG 364
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
DB 365 CGCAGAGAGAGCGCGGGGACTCGACCTCGCGCTGACCCAGCTGCTCATGGCCTCTCCTC 424
QY 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
DB 425 GAGGACGCGCAGATCGCGGACTGCTCAAGAGAGCGCGCGCTGCTCGCGCTCGCGGGTGC 484
QY 139 SerGluValGluLysLeuArgGlyLysGlu---GlyLysLysValGluSerAlaSerGly 157
DB 485 GCCGAGCTCGAAGCTCCGCGCGGGGACAACCTCGCGAAGGTTCGAGTCCGCTTCCTCGGC 544
QY 158 AspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLys 177
DB 545 GACACCACTTCGAGGCGCTCAAGACGTACGCGCGCGACCTCGTCGAGGTGGCGGCAAG 604
QY 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
DB 605 CTCGACCCCGTTCATCGCGCGCGCAGAGAGATCCGCGCGCTCGTGGCGCATCTCTCGCGC 664
QY 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
DB 665 CGCACCAAGAACACCCCGCTCTCATCGCGGAGCGCGCGCTGCGCAAGCGCGCTGCTC 724
QY 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
DB 725 GAGGCGCTCGCGCAGCGCGTCTCGCGGGGAGCGTCCCGAGCAACCTGCTCGACGTCGCG 784

```

238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
 785 CTCGTCGCGTCGACATAGGCGCGCTCGTCCGCGCCCAAGTACCGCGCGAGTTCGAG 844
 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 845 GAGCGGCTCAGGCGGCTGCTCAGGAGGTGAGGAGCGCGACGCGCAAGGTTCATCTCTC 904
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 905 ATCAGCAGATACACCTCTCTCGCGCGCGGAGGAGCGGAGGCGCTCATGGACGCGGCC 964
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
 965 AACCTCTTCAAGCCCATGCTCGCGAGGCGCGCTCGGTCGATCGCGCGCCACCGAGCTG 1024
 318 GluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 337
 1025 GAGGATACCGCAAGTACGTGAGAGGAGCGCGCTTCGAGCGACGCTTCACAGAGGTG 1084
 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
 1085 TACGTGGCGGAGCGCGCTCGCGACACCATCAGCATCTCTCGGAGGCGCTCAAGGAGAG 1144
 358 TyrGluGlyHisHisGlyValArgTleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 1145 TACGAGGCGGACCGCGGCTCGCATCCAGGACCGCGCATCTGCTCGCGCGCGAGCTC 1204
 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 1205 TCGGCAGGTATCATGTCGCGCGCCACCTCGCGGACCAAGGCGCATCGACTGTGGACGAG 1264
 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 1265 CGGTGCGGCAACGTGAGGCTGAGTGGACAGCCAGCGCGGAGGATCGCAATCTGGAG 1324
 418 ArgLysArgMetGlnLeuGluIleGlnLeuHisAlaLeuGluArgGlyLysAspLysAla 437
 1325 CGGAGCGGATTCAGCTGAGTGGAGCTGACCGCGCTGGAGAGGAGAGGACAGGAGGCC 1384
 438 SerLysAlaArgLeuIleGlnValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 1385 AGCAAGGCGCGCTGCTGGAGTGGAGGAGGCTGGACGACCTGAGGAGCAACAGTGCAG 1444
 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 1445 CCGCTTCACTCAAGTACAGGAGGAGGAGGAGGAGGATCGACGAGATCCGGAAGCTGAAG 1504
 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
 1505 CAGCGCGGAGGAGGCTGAGTTCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1564
 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 1565 CGGCTGCGGAGCTCAAGTACGCGCGCTTCCAGGAGATCGACGCGCGCTTCCGGAAGCTG 1624
 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 1625 GAGGCGGAGACCGGCG---GAGAACCTGATGCTCACCGAGACCGTCCGCGCGCGAGCATC 1681
 538 AlaGluValValSerArgThrThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 1682 CGGAGGTGTGAGCGGTGGAGCGGCTATCCCGGCTATCCCGGCTCGGCGGAGGAGGAGGAG 1741
 558 GluArgLeuIleGlyLeuAlaAspArgArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 1742 GAGCGGCTGCTGGCATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1801
 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 1802 GTGAAGCGGCTGCGCGGCGGCTGCTGAGGTCCAGGCGGCGGCTGCGGCGGCGGAGGAG 1861

598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
 1862 CCCACCGGCTGCT 1921
 618 AlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
 1922 GCCTCTCGCGAGCAGCTGTTCGACGACGAGAACCTGCTCTCTCTCTCTCTCTCTCTCTCT 1981
 638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis 657
 1982 TACATGAGCAGCATCT 2041
 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 2042 GAGGAGGCGGCGGCTGACGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2101
 678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 2102 GACGAGGTGGAGAGGCGCAGCTGCGGCTGTTCAACACCTCTCTCTCTCTCTCTCTCTCTCT 2161
 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
 2162 GCGCGGCTGACCGAGCGGCGAGCGGCGGAGCTTCCGGAACACCTGATCATCATG 2221
 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
 2222 ACCTCAACCTCGCGCGCGGAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2281
 738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
 2282 GTCGCGCGGAGCT 2341
 758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
 2342 CGGCTGAGCAGAGTGTGTCT 2401
 778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
 2402 CGCTTCAGATGAAGACGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2461
 798 ThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArg 817
 2462 ACCGAGCGCGCTGAGCAGCT 2521
 818 ProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArg 837
 2522 CCATCAGGAGGTGTGTCT 2581
 838 GluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAsp---LeuVal 856
 2582 GAGGAGATCGACGAGAACTGCTGTGTATCTGATCTGATCTGATCTGATCTGATCTGATCTG 2641
 857 TyrArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
 2642 TACGCGGTGAGCAGCGCGGCTGT 2701
 876 LeuIleHisIleAlaLeuGlyPro-----LysArgSerAspAlaAlaGlnAla 891
 2702 CT 2761
 892 ValLysLysMetArgIleGluGluIleGluAspAspAsnGluGluMetIleGluAsp 911
 2762 GTGAAGAAGATGAGGCTC-----ATGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAT 2815

RESULT 13
 AC130611/c
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) chromosome 5 clone P0483D07,
 *** SEQUENCING IN PROGRESS ***, 4 ordered pieces.
 ACCESSION
 AC130611.1 GI:22212962
 VERSION
 HTG; HTGS PHASE2.
 KEYWORDS
 Oryza sativa (japonica cultivar-group)

* are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 33927: contig of 33926 bp in length
 * 34026: gap of unknown length
 * 34027: 98487: contig of 64461 bp in length
 * 98488: 98587: gap of unknown length
 * 98588: 114278: contig of 15691 bp in length
 * 114279: 114378: gap of unknown length
 * 114379: 123438: contig of 9060 bp in length
 * 123439: 123538: gap of unknown length
 * 123539: 147456: contig of 23918 bp in length
 * 147457: 147556: gap of unknown length
 * 147557: 162429: contig of 14873 bp in length.

Location/Qualifiers

1. 162429
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="5"
 /clone="P0599F04"
 45748 a 34689 c 34184 g 47308 t 500 others

Alignment Scores:

d. No.: 2.63e-156 Length: 162429
 3691.00 Matches: 765
 81.00% Conservativity: 75
 73.77% Mismatches: 66
 81.12% Indels: 132
 2 Gaps: 9

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 120111 ATGAATCCGAGCAACTTCACGACACAGCAAGCAAGCGGCGCTCGTGGCGGCGCAGCAGATA 120052

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 120051 GCGTCGAGAGCGGGGACGCGGAGCTGACGCGCTGCACCTGTTGGCGCGCTGGCGGG 119992

41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
 119991 GACAAGGGCGGATATCTCCGACAGCGGATCTCGACGGCTCCGCGGGGACGCGCGCGG 119932

60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLeuProSerGlnSerPro 78
 119931 CCGGACTCGTTCGAGCGCGTGGTGGAGCGGGCGGCTCAAGAAGTGGCCCTCGCAGTCGCG 119872

79 ProProAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAla 98
 119871 CCGCGGACAGCGTGGCGGGTCCACGCGCTGATCAAGTTCATCCGCGCGCGCAGTCC 119812

99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 119811 GCGCAGAAGAAGCGCGGACTCCACCTCGCGCTCGACGAGTGTCTGTCGGCGCTCTG 119752

119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 119751 GAGACTCGCTAATCTCCGACTGCCCTCAAGGAGCGCGGCTGTCCCGCGCGGGTGC 119692

139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 119691 GCGGAGCTCGAAGACTCCGTGGCGGGAGGGCGCAGAGTGGAGTCCGCGTCCGCGGAC 119632

159 ThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLysLeu 178
 11952 ACCTGCGAGGAGCGCGCGGAGTGGACTCTCGCGCGGCTCGCCGACCTCAAGTACGGC 118493

Db 119631 ACCAACTTCAGGCGCTCAAGACTAGCGCGCGAGCCTCGTCGAGCAGCGCGCTAAAGCTC 119572

Qy 179 AspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgArg 198

Db 119571 GACCCCTCATTTGGCGCGAGAGATCCGCGCGTGTGTGCGATCTCTGTGCGCGCGC 119512

Qy 199 ThrLysAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218

Db 119511 ACGAAGAACCAACCCGCTACTCATCGCGGAGCCCGCGTGGCAAGACGCGCTGTGGAG 119452

Qy 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238

Db 119451 GGCCTCGCAGAGCATCTGTCGCGGGAGCTGCCAGCAACCTGTCTCGACGCTCCGCTC 119392

Qy 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyValLysThrArgGlyGluPheGluGlu 258

Db 119391 ATCGGCTCGACATGGCGCGCTGTCGCGCGGCGCAAGTACCGCGCGAGTTCGAGAGAA 119332

Qy 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278

Db 119331 CGGCTCAAGCGCGTCTCAAGGAGGTGGAGGCGGAGGCAAGGTGATCTCTTCATC 119272

Qy 279 AspGlnIleHisLeuValLeuGlyAlaGlyLysThrGluClySerMetAspAlaAlaAsn 298

Db 119271 GACGAGATACACCTCTGCTCGCGCGCGGAGGAGCGAGGCTCCATGGACGCGGCAAC 119212

Qy 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318

Db 119211 CTGTTCAAGCCATCTCGCGGCGGCGCAGCTCGGTGCTATCGCGCCACGACGCTGGAG 119152

Qy 319 GluTyrrArgLysTyrrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrr 338

Db 119151 GAGTACCGCAAGTACGTCGAGAGGAGCGCGCTTCGAGAGGCGGTTCACAGCAGGTGTT 119092

Qy 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyTyrr 358

Db 119091 GTCCGAGCGGAGCGTCCCGGACACCATCAGATCTCGGAGGCTCAAGGAGAGTAC 119032

Qy 359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378

Db 119031 GAGGGGACCAACGCTGTGAGATCCAGACCGCGCTGCTGCTCGCGCGCAGCTCTCG 118972

Qy 379 AlaArgTyrrIleThr----- 383

Db 118971 GCGAGGTATCATCAT-GGGTAAGTTGCTCTCGCGCTCTTCCATTCGCAATTCGCAACCTA 118913

Qy 384 -----GlyArg 385

Db 118912 CATGTCAAGCTCGAGTTCAATTTGTGCGCAATTTCTGCTCTGTAAATTTTCAGGACGC 118853

Qy 386 HisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGln 405

Db 118852 CACCTGCTGACAAAGCATGCGACCTGTCGAGAGGCTTCGCGAAGCTGTGAGGGTGCAG 118793

Qy 406 LeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIle 425

Db 118792 CTGACACCGCAGCGCAGAGAGATCGAACCTCGAGAGGAGGATTCAGCTGGAGGTG 118733

Qy 426 GluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuIleGluVal 445

Db 118732 GAGCACCAGCGCTGGAG 118673

Qy 446 ArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrrArgLys 465

Db 118672 AAGAAAGAGCTGGACGATCTGAGGACAAAGCTGAGCAATTCAGCATGAAGTACAGAGAG 118613

Qy 466 GluLysGluArgIleAspGluIleArgArgLysGlnLysArgGluGluLeuMetPhe 485

Db 118612 GAGAAGGAGAGAGATCGACGAGATCAGAGAGCTGAAGCAGCGCGCGAGGAGCTCCAGTTC 118553

Qy 486 SerLeuGlnGluAlaGluArgArgTyrrAspLeuAlaArgAlaAlaAspLeuArgTyrrGly 505

Db 118552 ACCTGCGAGGAGCGCGGAGTGGACTCTCGCGCGGCTCGCCGACCTCAAGTACGGC 118493

Qy	808	userTyAspPro	812
Db	117415	GTCTATGATCCGGTACCGCAGCATCACTACAGTCTAGGTCGAAGATTTTTCACAAAGAAATT	117356
Qy	813	-----ValTyrGlyA	816
Db	117355	CTCGTCAATTGATTCTTAACCTCTCTCTGACTCTGCTACTACTGACAGGTTTATGGCG	117296
Qy	816	laArgProIleArgArgTirpMetGluLysValValThrGluLeuSerLysMetValV	836
Db	117295	CGAGCGCATCAGAGGTGGTGCAGGAAGAGGGTGGTGCAGCAGCTGTCTCAAGATGCTGA	117236
Qy	836	alArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGly	854
Db	117235	TCCAGGAGAGATCGACGAGAACTCGCGGTGTATCATCGACGCGCGCCGACAGGACG	117176
Qy	854	spLeuValTyrArgValGluSer---GlyGlyLeuValAspAlaSerThrGlyLysLys	873
Db	117175	AGCTGGCTCAGAGGTGGACAAACCGCGGGGGTGGTGAACCGCGAGACGGCGCAGAGT	117116
Qy	873	erAspValLeuIleHisIleAlaAsnGlyProLysArg---SerAspAlaAlaGlnAlav	892
Db	117115	CGGACATCTCTCATCCAGGTGCCAACCGCGCGCCACCGGAGGACGCGCGCAGGCGG	117056
Qy	892	allLysLysMetArgIleGluIleGluAspAspAspAsnGluGlu	907
Db	117055	TGAAGAGATGAGATCATGAGGACGAGGACGCGCATGCGACGAGAG	117009
RESULT 15			
LOCUS	OSA316025	3109 bp DNA linear	PLN 27-AUG-2002
DEFINITION	Oryza sativa hsp101 gene for heat shock protein 101, exons 1-5.		
ACCESSION	AJ316025		
VERSION	AJ316025.1	GI:22535405	
KEYWORDS	heat shock protein 101; hsp101 gene.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.		
AUTHORS	Agarwal, M., Sahi, C., Young, T., Agarwal, S., Mitra, S., Ganeshan, K., Katiyar-Agarwal, S., Gallie, D.R. and Grover, A.		
JOURNAL	Isolation and analysis of HSP101 from rice		
REFERENCE	Unpublished		
AUTHORS	Agarwal, M.		
JOURNAL	Direct Submission		
FEATURES	Submitted (28-AUG-2001) Agarwal M., Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, Dhaula Kuan, New Delhi, Delhi-110021, INDIA		
source	Location/Qualifiers		
gene	1. 3109		
CDS	/organism="Oryza sativa (japonica cultivar-group)"		
	/mol_type="genomic DNA"		
	/cultivar="Nipponbare"		
	/db_xref="taxon:39947"		
	3. 3108		
	/gene="hsp101"		
	join(3..1159,1258..2061,2150..2431,2516..2710,2808..3108)		
	/gene="hsp101"		
	/product="heat shock protein 101"		
	/protein_id="CAC87117.1"		
	/db_xref="GI:22535406"		
	/translation="MNPDMFTKNEALVAAEHTASEAGHAQLTFLHLVLAALAKGGILRAQISAQSGDAGAPDSFEVVSGALKKLPSQPPDSVPASTALIKVIRRAQSAQKRGDGLHADVQLLGLLESLSDCLKEAGVSARVRAELEKLRGEGEKVESASGD TNFQALTYGRDLVQEGKLDVFGRIEIRVRVRLISRRTKNPNVLIGPEVGKTAIV EGAQRIVRGDVFNLLDVLRIALDMGALVAGKYRGEFEERLKAIVKIVEEAEKGV ILFDIEHLVAGRTSGSDAANLFPMLARGQRCITGATTEEYRYKTYEKRHLPAER FQCVFAEYPSVPTTISIRGLKEGHHGVRODRALVVAALQSLARYIMGRHLPOK AIDLVGAGANVRVOLDISOPEDTNIERRKTOLVEFHALLFKKDKKASKARLVKVEKE		

LDDLRKLOPLTKYKKEKERIDEIRKLKORREELQFTLQEAERRMDLARVADLKVGA
 LQEIADVAKLSEETGENIMLTETVPEQIAEIVSRWTGIPVTRLGQNDERLVLGLAD
 RHQVRVGAQAVSAVAEVRSLRAGLRPOQPTGSLFPLGPTGVGKTELAKALAEQL
 FDENLRLIDMSYMEQHSVARLIGAPGYVHGEGGLTEQVRRRYPYSVILYDEVE
 KAHVAFNTLLQVLDGRLTDGQRTVDPRNTVIMTNSLGAELHLLAGMVKMSKVA
 ROLVQEVRRHPRPELLNRLDEIVLFDPLSHEQLKVARLQWKDVAVRLAGRGVALAV
 TDAALDVILSLSDVPYVGARPIRRWIEKRVVTQLSKMLIOBEIDENCVTYIDAAAPHKO
 ELAYVRDRGGVLVNAETGQKSDILLIQVFNPAATGSDAAQAVKWKRIEDEDEGMDDEE"
 <3..1159

exon
 /gene="hsp101"
 /number=1

intron
 /gene="hsp101"
 /number=1

exon
 /gene="hsp101"
 /number=2

intron
 /gene="hsp101"
 /number=2

exon
 /gene="hsp101"
 /number=3

intron
 /gene="hsp101"
 /number=3

exon
 /gene="hsp101"
 /number=4

intron
 /gene="hsp101"
 /number=4

exon
 /gene="hsp101"
 /number=5

BE COUNT 617 a 951 c 1048 g 493 t
 LGIN

Alignment Scores:
 Seq. No.: 1,24e-156 Length: 3109
 Matches: 760
 Conservative: 77
 Mismatches: 69
 Indels: 132
 Gaps: 9

us-09-812-350-17 (1-911) x OSA316025 (1-3109)

- 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
- 3 ATGAATCCGACAACTTACGACACAGACCAACGAGGCGCTCGTGGCGCGCACAGAGATA 62
- 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
- 63 GCCTCGAGCGCGGCGCACGCGAGCTGACGCGCGCTGACCTGCTGGCGCGCTGGCGCG 122
- 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
- 123 GACAAAGCGCGGATCTCCGAGAGGAGTCTCCAGCGCTCGCGCGGACGCGCGCGG 182
- 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
- 183 CCGGACTCGTTGAGCGCGTGGTGAGCGGCGCTCAAGAGCTGCGCTCGAGTGGCGG 242
- 79 ProProAspIleProAlaSerSerSerLeuIleIleValIleArgAlaGlnAla 98
- 243 CCGCGGACAGCTGCGCGCTCATCAAGTCTATCCGCGCGCGCGAGTCC 302
- 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeu 118
- 303 CGCGAGAGAGCGCGGAGTCTCCACCTCGCGTGCACAGCTGCTGTCGCGCTCTG 362
- 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138

363 GAGGACTCGCTAATCTCCGACTCCCTCAAGGAGCGCGCGTGTCCGCGCGCGGTGCGC 422
 139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 423 CCGGAGTCCAGAGTCCGTTGGCGGAGCGCGCGAGGTGAGTCCGCGTCCGCGGAC 482
 159 ThrAsnPheGlnAlaLeuLysThrTyTGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
 483 ACCAACTTCCAGCGCTCAAGACTACCGCGCGAGCTCTGTGCGCATCTGTGCGCGCGC 542
 179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 198
 543 GACCCGCTCATTTGGCGCGACGAGAGATCCGCGCGCTGTGCGCATCTGTGCGCGCGC 602
 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGlu 218
 603 ACAGAGAACCAACCCGCTACTCATCGCGAGCGCGCGTTCGCGAAGACGCGCGTGGAG 662
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 663 GGCCTCGCGACGCGCTCTGTGCGGCGGACGTGCCAGCAACTGCTCGAGCGTCCGCTC 722
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGluPheGlu 258
 723 ATCGCGCTCGACATGGCGCGCTCTGTGCGGCGCGCAAGTACCGCGCGAGTTCGAGGA 782
 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
 783 CGGCTCAAGCGCGTCTCAAGGAGGTGGAGAGCGCGCGGAGGAGTGTATCTCTTCATC 842
 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGlySerMetAspAlaAlaAsn 298
 843 GACGAGATACACCTCGTCTCGCGCGCGGAGGACGAGGCGTCCATGAGCGCGCGCAAC 902
 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 903 CTGTTCAAGCAATGCTCGCGAGGCGCGCTCGGTGTGTCATCGCGCGCACGCGTGGAG 962
 319 GluTyArgLysTyValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTy 338
 963 GAGTACCGAGTACGTGAGAGGACCGCGTTCGAGAGGCGGTTCAGAGGAGTTC 1022
 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysTy 358
 1023 GTCCGCGAGCGCGTCCGCGACCATCAGCATCTCGAGGCGCTCAAGGAGAGTAC 1082
 359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 1083 GAGGCGCACCAACGCTGTGAGGATCCAGACCGCGCGCTGTGTCGCGCGCGAGCTCTCG 1142
 379 AlaArgTyIleThr--- 383
 1143 CCGAGGTACATCAT-GGGTAAGTTGCTCTTCCGCTCTTCCCATTTGCCATTGCCAACCTA 1201
 384 ---GlyArg 385
 1202 CATTTCAAGCTCGAGTTCATTTGTGCGCAATTTCTGCTGTATTTCTTCAGGACGC 1261
 386 HisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGln 405
 1262 CACTGTGCTGCAAGCGATCGACCTCGTGGGAGCGCGCGAGAACGTGAGGTGCGAG 1321
 406 LeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIle 425
 1322 CTGACAGCCAGCAGGAGATCGACAACTCGAGAGGAGAGGATTACGTGGAGGTG 1381
 426 GluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuGluVal 445
 1382 GAGCACCAACGCGCTCGAGAGGAGAGGACAAAGCAAGCAAGCGCGCTGTGCGAGT 1441
 446 ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyArgLys 465

```
1442 AAGAAAGAGCTGGACGATCTGAGGACAAAGCTGCAGCCATTGACGATGAAGTACAGGAAG 1501
466 GluIysGluArgIleAspGluIleArgArgLeuIysGlnIysArgGluGluLeuMetPhe 485
1502 GAGAGGAGAGGATCGACGAGATCAGGAAGCTGAAGCAGCGCGGAGGAGCTCCAGTTC 1561
486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArgfyrGly 505
1562 ACGTGCAGGAGGCGGAGCGCGGATGACCTCGCCCGCTCGCCGACCTCAAGTACGGC 1621
506 AlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsn 525
1622 GCGTCCAGGAGATCGACGTCGCCATCGCAAGCTCCAG---AGCGAGACGGGGAGAAC 1678
526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThr 545
1679 CTGATGCTACCGAGACCGTTCGGCCCGCAGCAGATCGCCGAGGTGGTGAGCCGCTGACC 1738
546 GlyIleProValThrArgLeuGlyGlnAsnGluIysGluArgLeuIleGlyLeuAlaAsp 565
1739 GGCAATCCCGGTGACCGGCTCGGCCAGAACGACGAGGAGGCTCGTGGCCCTCGCGGAC 1798
566 ArgLeuHisIysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
1799 AGGCTGCACACGAGGGTGGTGGCCAGCGGAGGCGGCTGAGCGCCGCTCGCCGAGCGGTG 1858
586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
1859 CTGAGGTTCGGCGCGGACATTGGCCGCGCGAGCGGCCACCGGCTCGTTCCTTCCTC 1918
606 GlyProThrGlyValGlyIysThrGluLeuAlaIysAlaLeuAlaGluGlnLeuPheAsp 625
1919 GGTCCCAACCGCGTCCGCAAGACCGAGCTCGCCAAAGCCCTCGCCGAGCAGCTGTTCCGAC 1978
626 AspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSer 645
1979 GACGAGAACCTGCTCGTCCGCATCGCATGTCCGAGTACATGAGCAGCAGCACTCCGTCGT 2038
646 ArgLeuIleGlyAlaProPro-Gly----- 653
2039 CGCCTCATCGAGACCACTCGGTGGTAGTATCAATCCCAATGTTGTTTTCACAGTTCCTG 2098
654 -----TyrValG1 656
2099 CAACCGAATTTCGGGTGTGACATGGGGAAATTTCTCTCTCTGAATGAAGCTACGTCGG 2158
656 yHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleIle 676
2159 CCATGAGGAAAGGTGGGAGCTGACGAGCAAGTCCGCGCGGCGGTACAGCGGTGATCCT 2218
676 uPheAspGluValGluIysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAs 696
2219 GTACGACGAGGTGAGAGGCGCACGTCGGGTGTTCAAACACTCTGCTCCAGGTTCCTGA 2278
696 pAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleI1 716
2279 CGACGGGAGGCTCACCGACGGGCAAGCGCGCACGCTGGACTTCAGGAACACCGGTGATCAT 2338
716 eMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyIysValThrMe 736
2339 CATGACCTCCAACTCGCGCGCGAGCACTCTCTCGCGCGCATGCTCGCGCAAGAACTCCAT 2398
736 tGluValAlaArgAspCysValMetArgGlu----- 746
2399 GAAGGTGCCCGTGCACCTTGTCATGAGGAGGTATGATGGCCCTTTGCTTCCACAAACC 2458
747 -----Valar 748
2459 ATGGATTCAATTGCAATGCGATGCTGAATTGATTTCGATATTTGTGCGAGGTGAG 2518
748 GlyHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLe 768
2519 GAGGCACTTCGCTCGAGTTCCTGAAACCGGCTCGACGAGATTGTGATATTCGATCCCT 2578
```

Search completed: February 13, 2004, 04:00:20
Job time : 7680 secs

```
QY 768 uSerHisaspGlnLeuArgIysValAlaArgGluGlnMetIysAspValAlaValArgLe 788
Db 2579 GTCCACAGCAACTCAGAGAGTCCCGCTCTCCAGATGAAGACGTCGCGCTCCGCT 2638
QY 788 uAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaG1 808
Db 2639 CGCCGCGAGGGGTGTCGCCCTGCGCGCTCACCGACCGCGCTGGACGTCAATTCCTCTCT 2698
QY 808 uSerTyrAspPro----- 812
Db 2699 GTCCCTATGATCCGCTACCGCAGCAGCATCAAATATACGTCGAAGTTTTTCAAAGAAATT 2758
QY 813 -----ValTyrG1YA 816
Db 2759 CTCGTCGAATTTGATTCTCTCTCTGTGACTCTGTACTACGTACAGGTTTATGGCG 2818
QY 816 laArgProIleArgArgTyrMetGluIysLysValValThrGluLeuSerIysMetValV 836
Db 2819 CGAGCCGATCAGGAGGTGGATCGAGAAAGAGGTGTCACGACGCTGTCCAAGATGCTGA 2878
QY 836 alArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGly-----AlaG1YA 854
Db 2879 TCCAGGAGGAGATCGACGAGAACTGCACAGTGTATCATCGCGCGCGCCGCCACAGGACG 2938
QY 854 spLeuValTyrArgValGluSer---GlyGlyLeuValAspAlaSerThrGlyIysLys 873
Db 2939 AGCTGGCTTACAGGTGGACCAACCGCGGGGGTGGTCAACCGCGAGACGGGCGCAGAAGT 2998
QY 873 erAspValLeuIleHisIleAlaAsnGlyProLysArg---SerAspAlaAlaGlnAlaV 892
Db 2999 CGGACATCTCATCCAGGTGCCCAACCGCGCGCGCCACCGGAGGCGACCGCGCGAGCGG 3058
QY 892 alIysLysMetArgIleGluGluIleGluAspAspAsnGluGlu 907
Db 3059 TGAGAAGATGAGGATCATGAGAGCAGAGGACGGCATGGAACAGAG 3105
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:31:27 ; Search time 11016 Seconds
(without alignments)
11530.900 Million cell updates/sec

le: US-09-812-350-30
fect score: 3105
quence: 1 aaagtatcaattttacac.....tctttgggttttttcaata 3105

ring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

urched: 2888711 seqs, 20454813386 residues
al number of hits satisfying chosen parameters: 5777422

imum DB seq length: 0
imum DB seq length: 2000000000
t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3105	100.0	3105	8	ATU13949	U13949 Arabidopsis
2	2732.8	88.0	2736	6	AX505540	AX505540 Sequence
3	2732.8	88.0	2736	6	AX589851	AX589851 Sequence
4	2535.8	81.7	6376	8	AF218796	AF218796 Arabidops
5	2532.2	81.6	50821	8	AC020579	AC020579 Arabidops
6	1638.4	52.8	3049	8	SOY8B100HS	L35272 Glycine max
7	1564.4	50.4	3052	8	AF083343	AF083343 Nicotiana
8	1542.4	49.7	1570	8	AY062596	AY062596 Arabidops
9	1345	43.3	3084	8	AF133840	AF133840 Zea mays
10	1334.2	43.0	3058	8	AF174433	AF174433 Triticum
11	1289.2	41.5	3148	8	AF332981	AF332981 Oryza sat
12	1246.8	40.2	3107	8	AF097363	AF097363 Triticum
13	1236.2	39.8	2821	8	AF083344	AF083344 Triticum
14	978.8	31.5	197976	8	ATCHRIV39	AL161539 Arabidops
15	978.8	31.5	206606	8	ATFCAL	Z97336 Arabidopsis
16	931.4	30.0	2045	8	AF083327	AF083327 Zea mays
17	896.4	28.9	145149	2	AC130611	AC130611 Oryza sat
18	896.4	28.9	162429	2	AC132491	AC132491 Oryza sat
19	888.4	28.6	3109	8	OSA316025	AJ316025 Oryza sat
20	807	26.0	3942	8	AF077337	AF077337 Zea mays
21	750.4	24.2	301150	1	AP003190	AP003190 Clostridi
22	706.4	22.8	10382	1	AE007611	AE007611 Clostridi
23	691.4	22.3	303671	1	AE015937	AE015937 Fusobacte
24	679.2	21.9	10700	1	AE010496	AE010496 Clostridi
25	677.6	21.8	342800	1	AP003598	AP003598 Nostoc sp
26	673.2	21.7	300547	1	AE016954	AE016954 Enterococ
27	668.2	21.5	43980	6	AX413014	AX413014 Sequence
28	668.2	21.5	258650	1	AL596171	AL596171 Listeria
29	668.2	21.5	349980	6	AX417047	AX417047 Sequence
30	666.8	21.5	2776	1	SSU20646	U20646 Synchococc
31	665.6	21.4	290029	1	AE017027	AE017027 Bacillus
32	659.8	21.2	11001	1	AE006383	AE006383 Lactococc
33	656.6	21.1	3247	8	AF449501	AF449501 Paracocci
34	656	21.1	300300	1	AP005373	AP005373 Thermosyn
35	655	21.1	300885	1	AE012559	AE012559 Xylella f
36	653.8	21.1	295050	1	AL591982	AL591982 Listeria
37	653.8	21.1	301442	1	AE016798	AE016798 Vibrio vu
38	653.8	21.1	349980	6	AX641671	AX641671 Sequence
39	652.8	21.0	304708	1	AE017001	AE017001 Bacillus
40	652	21.0	282610	1	RPXX01	AJ235270 Rickettsi
41	651.8	21.0	10092	1	AE003889	AE003889 Xylella f
42	651.6	21.0	2863	1	AF061279	AF061279 Plectonem
43	650.2	20.9	4109	1	AF016634	AF016634 Lactococc
44	649.4	20.9	13313	1	U32767	U32767 Haemophilus
45	649.4	20.9	110000	6	AR274513_08	Continuation (9 of

ALIGNMENTS

RESULT 1
ATU13949
LOCUS Arabidopsis thaliana heat shock protein AtHSP101 mRNA linear PLN 06-JUN-1995
DEFINITION complete cds.
ACCESSION U13949
VERSION U13949.1 GI:537445
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 3105)
Schirmer,E.C., Lindquist,S. and Vierling,E.

TITLE
 JOURNAL
 EDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 TURES
 source

An Arabidopsis heat shock protein complements a thermotolerance defect in Yeast
 Plant Cell 6 (12), 1899-1909 (1994)
 95170291
 7866032
 2 (bases 1 to 3105)
 Direct Submission
 Submitted (24-AUG-1994) Elizabeth Vierling, Biochemistry,
 University of Arizona, Life Sciences South, Tucson, AZ 85721, USA
 Location/Qualifiers
 1..3105
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strains="Columbia"
 /db_xref="taxon:3702"
 /clone="1-18"
 /tissue_type="leaf"
 /clone_lib="lambda ZAP (Stratagene) cDNA library from heat shock poly(A) RNA; K.W. Helm and E. Vierling, Nuc. Acids Res., 17:7995 (1989)"
 /dev_stage="adult"
 1..3105
 /gene="Athsp101"
 /note="This protein is a member of the highly conserved HSP100/clp family of proteins, and like other HSP100/clp proteins, Athsp101 has two consensus ATP-binding domains"
 /codon_start=1
 /product="Athsp101"
 /protein_id="AAA67927.1"
 /db_xref="GI:537446"
 /translation="MNPEKFTKNTETATHELAVNAGHAQFPLHLAAGLISDPTG
 IFPQALSAGENAAQSAERVINQALKLPQSPPPDIPASSLSLIKVIRAQAQKS
 RGDTHLVQDLIMGLLEDSQIRKLINEVGVATRVKSEVLEKRGKGVSSAGDYN
 FOALATYGRDLEQAGKLPDVGREDEIRRVVRLISRTKNPVLIGSPGVKTAIVE
 GLAQIRIVKGDVNSITDVLRLISLDMGLVAGAKYGFEEFLKSLVKEVEDAEGKVL
 FDEHLVGLDNGSTGMDAANLFPMLAGQLRCIGATTLEEVKVKYKEDKAAFERF
 QQVVAERSVPTIISLGLKEKYESHGVRIQDALINAAQLSARVITGBHLPKAI
 DLVDEACANRVQLDSQPEIDNLERKMQLELHALERKDRKASKARLIEVKELD
 DLRLKLPQTMKRYKERIDEIRLQKRELMFSLQEAERDYDLARAADLRVGAIQ
 EVESIALQEGTSSENWMLTENVGPEHIAEVSWSWTGIPVTRLGONEKERLIGLADR
 LHKRVQGNQAVNAVSEALILSRAGLGRAQPTGSLFLGPTGVGKTELAKALAEQLF
 DHENLVRLDMSEVMEQHSVSLICAPPGYVGHSEGGOLTRAVRRPYCVILFDEVEK
 AHVAVENTLIVLDGRGLTDGQGVTDVFNPSVLIINTSLGAHLLAGLTGKVTWEVAR
 DCVMREVRKHFPELNLRLDELIVDFDPLSHDQLRKARLQKMDVNRVLAERGVAVLT
 DAALDYILLASDYPVYGARPIKRWKVKVTELSKGVVREIDENSTVYIDAGAGDLV
 YRVESGLVDASTGKSDVLIHANGPKRSDAQAQVKKRIEIEDDDDDNEEMIED"
 2899..>3105
 /gene="Athsp101"
 3105
 polyA_site
 /gene="Athsp101"
 /note="19 A nucleotides"
 891 a 541 c 852 g 821 t

Query Match 100.0%; Score 3105; DB 8; Length 3105;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAGTTATCAATTTTACACATTACCGCTATATCTGCTTGATTCTCTGCAAAAAGAGAA 60
 1 AAAGTTATCAATTTTACACATTACCGCTATATCTGCTTGATTCTCTGCAAAAAGAGAA 60
 61 GACTTTTACCGAGAGAGAGTCTCTGGTCTATTGAGAAATCTCAACGAAACAAACCCAG 120
 61 GACTTTTACCGAGAGAGAGTCTCTGGTCTATTGAGAAATCTCAACGAAACAAACCCAG 120
 121 TTCTCATATATCGTTTAAAGGGATTACAAAAGCTAATCGAAGATGATCCAGAGAAATTC 180

121 TTCTCATATATCGTTTAAAGGGATTACAAAAGCTAATCGAAGATGATCCAGAGAAATTC 180
 181 ACACAAGAGACAAACGAGACAAATTCCTACAGCTCATGAGCTAGCTGTAATGACGACAT 240
 181 ACACAAGAGACAAACGAGACAAATTCCTACAGCTCATGAGCTAGCTGTAATGACGACAT 240
 241 GCTCAATTCATCTCTTTGCAATTTAGCTGGTCTTTGATCTCTGATCCACCGGTATATTT 300
 241 GCTCAATTCATCTCTTTGCAATTTAGCTGGTCTTTGATCTCTGATCCACCGGTATATTT 300
 301 CCTCAGCATCTCTAGTCCGCTGCGGAGACGACGCTCAATCTCTGTAAGAGAGATC 360
 301 CCTCAGCATCTCTAGTCCGCTGCGGAGACGACGCTCAATCTCTGTAAGAGAGATC 360
 361 AATCAAGCCTTGAAGAAGCTTCTCTCACAAATCTCTCCACCTGATGATATCCAGCAGT 420
 361 AATCAAGCCTTGAAGAAGCTTCTCTCACAAATCTCTCCACCTGATGATATCCAGCAGT 420
 421 TCTAGTCTTATTAAGTCTATCTGCTGCTCAGCTGCTCAGAGTCAAGAGTGAATCT 480
 421 TCTAGTCTTATTAAGTCTATCTGCTGCTCAGCTGCTCAGAGTCAAGAGTGAATCT 480
 481 CATTTGCTCTTCAACCAAGTTGATTATGGGTCTCTTGAAGATTTCTCAAAATCAGGAGTTG 540
 481 CATTTGCTCTTCAACCAAGTTGATTATGGGTCTCTTGAAGATTTCTCAAAATCAGGAGTTG 540
 541 TTGAACGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 541 TTGAACGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 601 AAAGAGGAGAGAGAGTCTGAGAGTCTTCAAGGGGACACAAATTTTCAAGCTTAAAGACT 660
 601 AAAGAGGAGAGAGAGTCTGAGAGTCTTCAAGGGGACACAAATTTTCAAGCTTAAAGACT 660
 661 TATGAGAGAGATTTGGTTGAGCAAGCAGGAGAGCTTGTCTGCTGATGCTGCTGCTGATGAG 720
 661 TATGAGAGAGATTTGGTTGAGCAAGCAGGAGAGCTTGTCTGCTGATGCTGCTGCTGATGAG 720
 721 GAGATTAGAGAGTCTGAGAGTCTTCAAGGGGACACAAATTTTCAAGCTTAAAGACT 780
 721 GAGATTAGAGAGTCTGAGAGTCTTCAAGGGGACACAAATTTTCAAGCTTAAAGACT 780
 781 GGAGAGCAGAGTCTGTTGAGCAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 781 GGAGAGCAGAGTCTGTTGAGCAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 841 GGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTGCTGACATGGCTGCTGCTG 900
 841 GGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTGCTGACATGGCTGCTGCTG 900
 901 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 901 GTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 961 GTTGAGAGCCTGAGGAGAGTCTTCTTTTATGAGAGATTTGATGAGATTTGCTGCTGCTG 1020
 961 GTTGAGAGCCTGAGGAGAGTCTTCTTTTATGAGAGATTTGATGAGATTTGCTGCTGCTG 1020
 1021 GCTGGCAAAAATCTGAGGAGTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 1021 GCTGGCAAAAATCTGAGGAGTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 1081 GGGCAGCTTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 1081 GGGCAGCTTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 1141 GATCTGCTCTTCTGAGGAGGTTCCACAGTCTATGTTGCGGAGCCTGCTGCTGCTGCTGCTG 1200
 1141 GATCTGCTCTTCTGAGGAGGTTCCACAGTCTATGTTGCGGAGCCTGCTGCTGCTGCTGCTG 1200
 1201 ACCATTAGTATCTCTTCTGAGGAGTCTCAAGGAGAGTATGAGGAGATCATGCTGCTGCTG 1260
 1201 ACCATTAGTATCTCTTCTGAGGAGTCTCAAGGAGAGTATGAGGAGATCATGCTGCTGCTG 1260

source

1. .2736

/organism="Arabidopsis thaliana"

/mo_type="genomic DNA"

/db_xref="taxon:3702"

GIN 784 a 481 c 785 g 686 t

query Match

88.0%; Score 2732.8; DB 6; Length 2736;

Best Local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;

atches 2734; Conservative 0;

163 ATGAATCCAGAGAAATTCACACACAGACAAACGAGACAAATGCTACAGCTCATGAGCTA 222

1 ATGAATCCAGAGAAATTCACACACAGACAAACGAGACAAATGCTACAGCTCATGAGCTA 60

223 GCTGTAATGAGGACATGCTCAATTCACCTCTTTCATTTAGCTGCTGCTTTCATCTCT 282

61 GCTGTAATGAGGACATGCTCAATTCACCTCTTTCATTTAGCTGCTGCTTTCATCTCT 120

283 GATCCACCGGTATATTTCTCAAGCAATCTTAGTGGCGGTGGCGAAGCGCAGCTCAA 342

121 GATCCACCGGTATATTTCTCAAGCAATCTTAGTGGCGGTGGCGAAGCGCAGCTCAA 180

343 TCTGCTGAAAGAGTCAATCAAGCCTTGAAGAGCTTCTTCAATCTCTCCACCT 402

181 TCTGCTGAAAGAGTCAATCAAGCCTTGAAGAGCTTCTTCAATCTCTCCACCT 240

403 GATGATATTCAGCGAGTCTAGTCTTATTAGGTCATTCGTCGCTCAAGCTGCTCAG 462

241 GATGATATTCAGCGAGTCTAGTCTTATTAGGTCATTCGTCGCTCAAGCTGCTCAG 300

463 AAGTCACAGAGTGATCAATTTGGCTGTGACCAAGTTGATTTAGGTCCTTTCAAGAT 522

301 AAGTCACAGAGTGATCAATTTGGCTGTGACCAAGTTGATTTAGGTCCTTTCAAGAT 360

523 TCTCAAAATCAGGATTTGTTGAACGAAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAG 582

361 TCTCAAAATCAGGATTTGTTGAACGAAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAG 420

583 GTTGAGAGCTTCGTGGGAAAGAGGAGAAAGTTGAGAGTCTTCAAGGGGACACAAAT 642

421 GTTGAGAGCTTCGTGGGAAAGAGGAGAAAGTTGAGAGTCTTCAAGGGGACACAAAT 480

643 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTGATCCT 702

481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTGATCCT 540

703 GTGATTTGCTGATGAGGAGATTAGAGAGTCTGAGGATTTCTTCAGGAGAACGAG 762

541 GTGATTTGCTGATGAGGAGATTAGAGAGTCTGAGGATTTCTTCAGGAGAACGAG 600

763 AACAAATCCTGCTTATTGGAGAGCCAGGAGTTGGTAAACAGCTGTGTTGAAGTTTA 822

601 AACAAATCCTGCTTATTGGAGAGCCAGGAGTTGGTAAACAGCTGTGTTGAAGTTTA 660

823 GCACAAAGGATTTGAAAGGAGATGTGCCCAACAGCTTTACTGATGTGAGATTAAATTCG 882

661 GCACAAAGGATTTGAAAGGAGATGTGCCCAACAGCTTTACTGATGTGAGATTAAATTCG 720

883 TTGGACATGGGTGCGTTAGTCTGCTGCTTAATACCGAGGAGAGTTTGAAGAAAGTTG 942

721 TTGGACATGGGTGCGTTAGTCTGCTGCTTAATACCGAGGAGAGTTTGAAGAAAGTTG 780

943 AAATCTGTTTGAAGAAAGTTGAGGAGCTGGAAGCAAGTGAATCTCTTTATTGATGAG 1002

781 AAATCTGTTTGAAGAAAGTTGAGGAGCTGGAAGCAAGTGAATCTCTTTATTGATGAG 840

1003 ATTCAATTTGGTCTTGGTCTGCTGCAAAATCTGAAGGTCGATGAGTCAAGCTTAATCTGTT 1062

841 ATTCAATTTGGTCTTGGTCTGCTGCAAAATCTGAAGGTCGATGAGTCAAGCTTAATCTGTT 900

1063 AAGCCATGTTAGCTAGAGGGCAGCTTCGATGATTTGGTCTCAACCGTTGAAGATAC 1122

901 AAGCCATGTTAGCTAGAGGGCAGCTTCGATGATTTGGTCTCAACCGCTTGAAGATAC 960

1123 AGGAATATTTGAGAAAGATGCTGCTTTGAGAGGAGGTTTCAACAGCTCTATGTTGG 1182

961 AGGAATATTTGAGAAAGATGCTGCTTTGAGAGGAGGTTTCAACAGCTCTATGTTGG 1020

1183 GAGCCAAAGTGTGCTGCTGACACCAATTAGTATCTTTAGAGGACTCAAGAGAGATGAGGGA 1242

1021 GAGCCAAAGTGTGCTGCTGACACCAATTAGTATCTTTAGAGGACTCAAGAGAGATGAGGGA 1080

1243 CATCATGTTGTTGGAATCCACAGACAGAGCTTTAAATGCTGCTCAGCTGTCTGCTCT 1302

1081 CATCATGTTGTTGGAATCCACAGACAGAGCTTTAAATGCTGCTCAGCTGTCTGCTCT 1140

1303 TACATAACTGTTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGG 1362

1141 TACATAACTGTTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGG 1200

1363 AATGTGAGAGTCCAGCTTGTATAGTCAACCTGAAGAGATTTGATAACCTTGAAGAGAGG 1422

1201 AATGTGAGAGTCCAGCTTGTATAGTCAACCTGAAGAGATTTGATAACCTTGAAGAGAGG 1260

1423 ATGCAGTGGAAATTTGAACTTTCAGCGCTTTGGAAGAGGAGAGATTAAGCCAGCAAGCT 1482

1261 ATGCAGTGGAAATTTGAACTTTCAGCGCTTTGGAAGAGGAGAGATTAAGCCAGCAAGCT 1320

1483 CGACTTATAGAGTGGCGAAAGAGCTTTGATGACCTGAGAGACAAAGCTTCAAGCTCTCACG 1542

1321 CGACTTATAGAGTGGCGAAAGAGCTTTGATGACCTGAGAGACAAAGCTTCAAGCTCTCACG 1380

1543 ATGAAATACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGGCTTAAACAGAAAAA 1602

1381 ATGAAATACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGGCTTAAACAGAAAAA 1440

1603 GAGAGCTCATGTTTCTTTGCGAGGAGGAGAGAGATTAAGCTTCAAGAGCTGCT 1662

1441 GAGAGCTCATGTTTCTTTGCGAGGAGGAGAGAGATTAAGCTTCAAGAGCTGCT 1500

1663 GATCTAAGATATCGCCCAATTCAGAAAGTGGAAATCTGCAATTTGCCCAAATTTGAAGGAA 1722

1501 GATCTAAGATATCGCCCAATTCAGAAAGTGGAAATCTGCAATTTGCCCAAATTTGAAGGAA 1560

1723 TCTTCTGAAGAGAAATGTGATGCTCAAGAAAAAGTTGGGCTGGAACACATTTGCTGAGGTT 1782

1561 TCTTCTGAAGAGAAATGTGATGCTCAAGAAAAAGTTGGGCTGGAACACATTTGCTGAGGTT 1620

1783 GTGAGCGTTGGACAGGATTTCCAGTGCAGAGACTTTGGCCCAAAATGAGAGAGAGAGGTTG 1842

1621 GTGAGCGTTGGACAGGATTTCCAGTGCAGAGACTTTGGCCCAAAATGAGAGAGAGAGGTTG 1680

1843 ATTGGTCTTCTGATAGTTGCAATAAGCGGTTTGGGACAGAAATCAAGCGGTAAATGCA 1902

1681 ATTGGTCTTCTGATAGTTGCAATAAGCGGTTTGGGACAGAAATCAAGCGGTAAATGCA 1740

1903 GTTTCGAGGCAATTTCTAAGGTCAGGGCAGGAGCTTTGGTAGGGCAACACAGCAACTGGA 1962

1741 GTTTCGAGGCAATTTCTAAGGTCAGGGCAGGAGCTTTGGTAGGGCAACACAGCAACTGGA 1800

1963 TCATTCTTATTTCTTGACCAACTGTTGTTGGCAAACTGAGCTGCCCAAGGCTTTGCT 2022

1801 TCATTCTTATTTCTTGACCAACTGTTGTTGGCAAACTGAGCTGCCCAAGGCTTTGCT 1860

2023 GAGCAGCTGTTGATGATGAAACCTCTTAGTTTCGGAATTTGATGATGCGGAATATATGGA 2082

1861 GAGCAGCTGTTGATGATGAAACCTCTTAGTTTCGGAATTTGATGATGCGGAATATATGGA 1920

2083 CAACACTCTGCTCTGCTGCTCATTTGGGACCAACAGGATGTTGGTTCACGAGGAGGT 2142

1921 CAACACTCTGCTCTGCTGCTCATTTGGGACCAACAGGATGTTGGTTCACGAGGAGGT 1980

2143 GGACAACTAACTGAGGCTGTGAGGAGCGGACCTTTATTTGTCATCTCTTTGATGAGGTTG 2202

1981 GGACAACTAACTGAGGCTGTGAGGAGCGGACCTTTATTTGTCATCTCTTTGATGAGGTTG 2040

ORIGIN

2203 GAGAGGCTCATGTTGCTGCTTCAACACTCTGCTCAAGTTTGGATGATGGTCGATTG 2262
2041 GAGAGGCTCATGTTGCTGCTTCAACACTCTGCTCAAGTTTGGATGATGGTCGATTG 2100
2263 ACAGCGGGCAAGCAGGACAGTGCATTTTCAGGAACCTCGGTGATATCATGACATCAAC 2322
2101 ACAGCGGGCAAGCAGGACAGTGCATTTTCAGGAACCTCGGTGATATCATGACATCAAC 2160
2323 CTTCGCTGAACACCTCTTTCAGGCTTCACTGGGAAAGTAACAATGGAAGTGGCCCGG 2382
2161 CTTCGCTGAACACCTCTTTCAGGCTTCACTGGGAAAGTAACAATGGAAGTGGCCCGG 2220
2383 GACTGTGTGATGCGGGAGGTGAGGAACACCTTCAGACCAGAGCTCTTGAACAGCGCTTGAC 2442
2221 GACTGTGTGATGCGGGAGGTGAGGAACACCTTCAGACCAGAGCTCTTGAACAGCGCTTGAC 2280
2443 GAGATTGTGTGCTGAGCCCTCTTTCAGATGACAGCTGAGGAAGTAGCTCGGCTTCAA 2502
2281 GAGATTGTGTGCTGAGCCCTCTTTCAGATGACAGCTGAGGAAGTAGCTCGGCTTCAA 2340
2503 ATGAAGACGCTGCTGCTGCGGCTTTCAGGAAGAGAGTTCCTTGGCAGTCACTGATGCT 2562
2341 ATGAAGACGCTGCTGCTGCGGCTTTCAGGAAGAGAGTTCCTTGGCAGTCACTGATGCT 2400
2563 GCTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGGTGTATGCTGCTAGGCCCTATAAGG 2622
2401 GCTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGGTGTATGCTGCTAGGCCCTATAAGG 2460
2623 AGATGGATGGAAGAAGGTGCTGACGAACCTGTCAGAGATGTTGTGCTGAGGAATC 2682
2461 AGATGGATGGAAGAAGGTGCTGACGAACCTGTCAGAGATGTTGTGCTGAGGAATC 2520
2683 GATGAACACTCCACTGTTTACATAGATGACGGCTGCTGATCTTGTGACCGGTTAGAA 2742
2521 GATGAACACTCCACTGTTTACATAGATGACGGCTGCTGATCTTGTGACCGGTTAGAA 2580
2743 AGTGAGGCTTAGTGAGCGCTTCAACAGGCAAGAGTCAAGTGTGCTGATTCATATTGCT 2802
2581 AGTGAGGCTTAGTGAGCGCTTCAACAGGCAAGAGTCAAGTGTGCTGATTCATATTGCT 2640
2803 AACGGCCAAAGAGAGTGTGACGCTGAGCGGTGAGAGATGAGGATCGAGGAATA 2862
2641 AACGGCCAAAGAGAGTGTGACGCTGAGCGGTGAGAGATGAGGATCGAGGAATA 2700
2863 GAAGATGACGATATGAGGAATGATCGAGATTAA 2898
2701 GAAGATGACGATATGAGGAATGATCGAGATTAA 2736

AX589851 2736 bp DNA linear PAT 24-JAN-2003
Sequence 33 from Patent WO02081695.
AX589851
AX589851.1 GI:27901086
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1. 2736
Zhu, T., Glazov, E. A., Meins, F., Wang, X. and Chang, H. S.
Genes that are modulated by posttranscriptional gene silencing
Patent: WO 02081695-A 33 17-OCT-2002;
Syngenta Participations AG (CH)
Location/Qualifiers
1. 2736
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
784 a 481 c 785 g 686 t

Query Match 88.0%; Score 2732.8; DB 6; Length 2736;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2734; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 163 ATGAATCCAGAGAAATTCACACAGCAACAAACGAGACAAATTTGCTACAGCTCATGAGCTA 222
DB 1 ATGAATCCAGAGAAATTCACACAGCAACAAACGAGACAAATTTGCTACAGCTCATGAGCTA 60
QY 223 GCTGTGAATGACGAGCACTGCTCAATTCATCTCTTTGCAATTTAGCTGGTGTGCTTGTATCT 282
DB 61 GCTGTGAATGACGAGCACTGCTCAATTCATCTCTTTGCAATTTAGCTGGTGTGCTTGTATCT 120
QY 283 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTCCGGTGGCGAGAAACGACGCTCAA 342
DB 121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTCCGGTGGCGAGAAACGACGCTCAA 180
QY 343 TCTGCTGAAAGAGTGATCAATCAAGCCTTTGAAGAAGTTCCTTCAATTCCTCTCCACT 402
DB 181 TCTGCTGAAAGAGTGATCAATCAAGCCTTTGAAGAAGTTCCTTCAATTCCTCTCCACT 240
QY 403 GATGATATTCAGCGAGTTCTAGTCTTATTAAGTTCATTCGTCGTCCTCAAGCTGCTCAG 462
DB 241 GATGATATTCAGCGAGTTCTAGTCTTATTAAGTTCATTCGTCGTCCTCAAGCTGCTCAG 300
QY 463 AAGTCAAGAGTGATCACTCATTTGGCTGTGACCAAGTTCATTTAGTGGTCTTCTTGAAGAT 522
DB 301 AAGTCAAGAGTGATCACTCATTTGGCTGTGACCAAGTTCATTTAGTGGTCTTCTTGAAGAT 360
QY 523 TCTCAAAATCAGGGAATTTGTTGAACGAAGTTCGGTGTAGCGAGCGGAGGTTAAAGTCTCAG 582
DB 361 TCTCAAAATCAGGGAATTTGTTGAACGAAGTTCGGTGTAGCGAGCGGAGGTTAAAGTCTCAG 420
QY 583 GTTGAGAACTCTGCGGGAACGAAGGAAAGTTCAGAGTTCCTTCAAGGAGCAACAAT 642
DB 421 GTTGAGAACTCTGCGGGAACGAAGGAAAGTTCAGAGTTCCTTCAAGGAGCAACAAT 480
QY 643 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTTGATCT 702
DB 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTTGATCT 540
QY 703 GTGATTTGCTGTGATGAGAGATTAAGAGATTCGCTGAGAGATTTCTTTTCGAGGAGAACGAG 762
DB 541 GTGATTTGCTGTGATGAGAGATTAAGAGATTCGCTGAGAGATTTCTTTTCGAGGAGAACGAG 600
QY 763 AACATCTCTGCTTATTTGGAGAGCCAGGTTGTTGTTAAACAGCTGTGTTGAAGGTTTA 822
DB 601 AACATCTCTGCTTATTTGGAGAGCCAGGTTGTTGTTAAACAGCTGTGTTGAAGGTTTA 660
QY 823 GCACAAAGGATTTGGAAGGAGATGTCGCCAACAGCTTTACTGTGATGAGATTAATTTTCG 882
DB 661 GCACAAAGGATTTGGAAGGAGATGTCGCCAACAGCTTTACTGTGATGAGATTAATTTTCG 720
QY 883 TTGGAATGCTGGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 942
DB 721 TTGGAATGCTGGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 943 AAATCTGTTTTCGAAGAAGTTCAGGAGCTGAGGCAAGGTAAGTTCCTTTTATTCATGATG 1002
DB 781 AAATCTGTTTTCGAAGAAGTTCAGGAGCTGAGGCAAGGTAAGTTCCTTTTATTCATGATG 840
QY 1003 ATTCAATTTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1062
DB 841 ATTCAATTTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 1063 AAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCTATGCTGCTAAGCCTTTGAGGAATAC 1122
DB 901 AAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCTATGCTGCTAAGCCTTTGAGGAATAC 960
QY 1123 AGGAATATGTTGAGGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1182
DB 961 AGGAATATGTTGAGGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020

1183 GAGCAAGTGTGCTGACACATTAAGTATCTTAGAGAGCTCAAGGAGAGTATGAGGGA 1242
 1021 GAGCAAGTGTGCTGACACATTAAGTATCTTAGAGAGCTCAAGGAGAGTATGAGGGA 1080
 1243 CATCATGTTGTCGGAATCAAGACAGAGCTCTTATAAATGCTGCTCAGCTCTGCTCGT 1302
 1081 CATCATGTTGTCGGAATCAAGACAGAGCTCTTATAAATGCTGCTCAGCTCTGCTCGT 1140
 1303 TACATPACTGTGTCGCAATTAACCGATAAAGCAATGATTTGTTGATGAGGCTTTGCGG 1362
 1141 TACATAAATGTTGTCGCAATTAACCGATAAAGCAATGATTTGTTGATGAGGCTTTGCGG 1200
 1363 ATGTGAGAGTCCAGCTGATAGTCAACCTGAAGAGATTTGATTAACCTTTGAAGGAGAGG 1422
 1201 AATGTGAGAGTCCAGCTGATAGTCAACCTGAAGAGATTTGATTAACCTTTGAAGGAGAGG 1260
 1423 ATGCAAGTGGAAATGAATTCACGCCCTTTGAAAGGAGAGATTAAGCCAGCAAGCT 1482
 1261 ATGCAAGTGGAAATGAATTCACGCCCTTTGAAAGGAGAGATTAAGCCAGCAAGCT 1320
 1483 CGACTTATAGAGTGGGAAAGAGCTTGATGACCTGAGAGCAAGCTTCAGCTCTCAGG 1542
 1321 CGACTTATAGAGTGGGAAAGAGCTTGATGACCTGAGAGCAAGCTTCAGCTCTCAGG 1380
 1543 ATGAATACAGAAAGGAGAAAGAGAGATTTGATGAGATTCGAAGGCTTTAAACAGAAAGA 1602
 1381 ATGAATACAGAAAGGAGAAAGAGAGATTTGATGAGATTCGAAGGCTTTAAACAGAAAGA 1440
 1603 GAAGAGCTCATGTTTTCTTTGAGAGGAGCAAGCAAGATATGACCTTTGCAAGAGCTGCT 1662
 1441 GAAGAGCTCATGTTTTCTTTGAGAGGAGCAAGCAAGATATGACCTTTGCAAGAGCTGCT 1500
 1663 GATCTAAGATATGGCCATTTCAAGAGTGGATCTGCAATTTGCCCACTTGAAGGACT 1722
 1501 GATCTAAGATATGGCCATTTCAAGAGTGGATCTGCAATTTGCCCACTTGAAGGACT 1560
 1723 TCTTCTGAAGAGATGTGATGCTCACAGAAAAGTTGGCCCTGGAACATTTGCTCAGGTT 1782
 1561 TCTTCTGAAGAGATGTGATGCTCACAGAAAAGTTGGCCCTGGAACATTTGCTCAGGTT 1620
 1783 GTGAGCGTTGAGAGGAGATTCAGTACAGACTTTGGCCAAATATGAGAGAGAGGTTG 1842
 1621 GTGAGCGTTGAGAGGAGATTCAGTACAGACTTTGGCCAAATATGAGAGAGAGGTTG 1680
 1843 ATTGCTCTGCTGATAGTTCATAGCGGTTGTGGACAGCAATCAAGCGTAAATGCA 1902
 1681 ATTGCTCTGCTGATAGTTCATAGCGGTTGTGGACAGCAATCAAGCGTAAATGCA 1740
 1903 GTTCTGAGGCAATTTCAAGGTCAGGCGAGGACTTTGGTGGGACCAACAGCCAACTGGA 1962
 1741 GTTCTGAGGCAATTTCAAGGTCAGGCGAGGACTTTGGTGGGACCAACAGCCAACTGGA 1800
 1963 TCAATCTTATCTTGGACCAACTGTTGTGGCAAACTGAGCTGCGCAAGCTCTTCT 2022
 1801 TCAATCTTATCTTGGACCAACTGTTGTGGCAAACTGAGCTGCGCAAGCTCTTCT 1860
 2023 GAGCAGCTGTTGATCATGAAACCTTTAGTTTCGATTGATATGTGCAATATATGGA 2082
 1861 GAGCAGCTGTTGATCATGAAACCTTTAGTTTCGATTGATATGTGCAATATATGGA 1920
 2083 CAACACTCTGCTCTCGCCCTCAATTTGGGGACCAACAGGAGTATGTTGGTCAAGGAGAGT 2142
 1921 CAACACTCTGCTCTCGCCCTCAATTTGGGGACCAACAGGAGTATGTTGGTCAAGGAGAGT 1980
 2143 GGACAACTACTGAGGCTGTGAGAGGAGCACTTATGTCATCTCTTGTATGAGTG 2202
 1981 GGACAACTACTGAGGCTGTGAGAGGAGCACTTATGTCATCTCTTGTATGAGTG 2040
 2203 GAGAAGCTCATGTTGCTCTTCAACACTCTGCTCCAAAGTTTGGATGATGCTGATG 2262
 2041 GAGAAGCTCATGTTGCTCTTCAACACTCTGCTCCAAAGTTTGGATGATGCTGATG 2100

2263 ACAGCGGCAAGCGAGGACAGTCGATTTTCAGGAATCTCGTGATATATCATCATCAAC 2322
 2101 ACAGCGGCAAGCGAGGACAGTCGATTTTCAGGAATCTCGTGATATATCATCATCAAC 2160
 2323 CTTTGTGCTGAAACACCTCTTTCAGGGCTAACTGGGAAAGTAAACAATGAAAGTGCCCGG 2382
 2161 CTTTGTGCTGAAACACCTCTTTCAGGGCTAACTGGGAAAGTAAACAATGAAAGTGCCCGG 2220
 2383 GACTGTGTATGCGGAGGTGAGGAACACTTTCAGACAGAGCTCTTGAACAGGCTTGAC 2442
 2221 GACTGTGTATGCGGAGGTGAGGAACACTTTCAGACAGAGCTCTTGAACAGGCTTGAC 2280
 2443 GAGATTGTGTGTTTCGACCCCTTTTCACATGACCACTTCAGGAAAGTAGCTCGGCTTCAA 2502
 2281 GAGATTGTGTGTTTCGACCCCTTTTCACATGACCACTTCAGGAAAGTAGCTCGGCTTCAA 2340
 2503 ATGAAAGACCTTCTGTCGGCTTGTGAAAGAGGAGTTGCTTTTGGCAGTCACTGATGCT 2562
 2341 ATGAAAGACCTTCTGTCGGCTTGTGAAAGAGGAGTTGCTTTTGGCAGTCACTGATGCT 2400
 2563 GCTTTGGACTATATCTTGGCAGAGGTTATGACCCGCTGATGCTAGGCTTAAAG 2622
 2401 GCTTTGGACTATATCTTGGCAGAGGTTATGACCCGCTGATGCTAGGCTTAAAG 2460
 2623 AGATGATGAGAGAGAGTGTGTCAGAACTGTCTCAAGATGCTTGTGCTGAGGAAATC 2682
 2461 AGATGATGAGAGAGAGTGTGTCAGAACTGTCTCAAGATGCTTGTGCTGAGGAAATC 2520
 2683 GATGAAACTCTCACTGTTTACATAGATGACGGCTGCTGATCTTGTACCGGTAGAA 2742
 2521 GATGAAACTCTCACTGTTTACATAGATGACGGCTGCTGATCTTGTACCGGTAGAA 2580
 2743 AGTGGAGTCTAGTGAAGCTTCAACAGGCAAGCTCAGATGCTGCTGATTTATATGCT 2802
 2581 AGTGGAGTCTAGTGAAGCTTCAACAGGCAAGCTCAGATGCTGCTGATTTATATGCT 2640
 2803 AACGGCCAAAGAGAGTGTGTCAGCTCAGGCGGTGAGAGATGAGGATTCGAGGAAATA 2862
 2641 AACGGCCAAAGAGAGTGTGTCAGCTCAGGCGGTGAGAGATGAGGATTCGAGGAAATA 2700
 2863 GAGATGACCATATGAGGAATGATCGAGGATTA 2898
 2701 GAGATGACCATATGAGGAATGATCGAGGATTA 2736

RESULT 4

AF218796
 LOCUS Arabidopsis thaliana 6376 bp DNA linear PLN 13-APR-2000
 DEFINITION Arabidopsis thaliana heat shock protein 101 (HSP101) gene, complete cds.

ACCESSION AF218796
 VERSION AF218796.1 GI:6715467

KEYWORDS

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE

1 (bases 1 to 6376)
 Hong, S.W. and Vierling, E.
 Mutants of Arabidopsis thaliana defective in the acquisition of
 tolerance to high temperature stress
 Proc. Natl. Acad. Sci. U.S.A. 97 (8), 4392-4397 (2000)

JOURNAL

MEDLINE 20226114

PUBMED 10760305

REFERENCE

2 (bases 1 to 6376)

Hong, S.-W. and Vierling, E.
 Direct Submission
 Submitted (22-DEC-1999) Biochemistry, University of Arizona, 1007
 E. Lowell St., Tucson, AZ 85721, USA

JOURNAL

FEATURES

1..6376
 Location/Qualifiers
 /organism="Arabidopsis thaliana"

```

/molecule="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="I"
/map="ngalli-ATATPASE"
<1132..>4279
/gene="HSP101"
/join(<1132..2281,2366..2547,2688..3313,3398..3875,
3980..>4279)
/gene="HSP101"
/product="heat shock protein 101"
/join(1132..2281,2366..2547,2688..3313,3398..3875,
3980..4279)
/gene="HSP101"
/note="clpB"
/codon_start=1
/product="heat shock protein 101"
/protein_id="AAF26423.1"
/db_xref="GI:6715468"
/translation="MNPBKFTHKNTIATHELAHVAGHAQFTPLHLAGALISDPTG
IPQAISSAGGENAQAQSRVQIDLANEVGVATARKVSEKLRGKGGKVESAGSDYN
RGDTHLVDQLIMGLLEDSQJRDLANEVGVATARKVSEKLRGKGGKVESAGSDYN
FOALTYGRDLVQAGKLDPIVGRDEIRRVRLISRTKNPVLIGPGVKTAIVE
GLAQRIVLGVDPNSLTDVLLSLDLMLAGVAGKYGEPERLKYKVEVEDAEKRVIL
FIDEHLVLGAGKTEGSDANLFPMLARQLRCIGATTLBYRYKYVEDAAFERF
QOYVAEPSPVDTISILRLAKEYEGHGVRIQDRALINAAQLSARYITGRHLPKAI
DLVDEACANVRVOLDSPREIDNLRKMOLEIELHALEREKQKARLLEVRKELD
DLRDLQPLTKMYRKEKRIEIRLQKREBELMFLQEAERYDLARAADLRGAIQ
EVESALIQLETSSENWMLTENUGPEHIAEVSVRWGTPVTRLOONERKILGLADR
LHKRVQGNQANVAESALRSRAGLRPQPTGSLFLGTPGVGKTELAKALAEQLF
DENLILVRIDMEYMEQHSVRLIGAPPGYHEEGQOLTEAVRRPYCVILFDEVRK
AHVAVENTLLQVLDGRITDQGRVDPFRNSVIIMTSLGAEHLGLAGLTKVTVMEVAR
DCMREVRKRPPELLNDELIVDPDPLSHDQKRVARLQMDKVAIRLAEGRVALAVT
DAALDYILAESVDPYVGRIRRMKKVVTLELSKVVREIIDEINSYVIDAGADLV
YRVESGLVDASTGKSDVLIHANGPKRSDRAQVKKWRIBIEDDDNEEMIED"
1914 a 1150 c 1470 g 1842 t
E COUNT
GIN
Query Match 81.7%; Score 2535.8; DB 8; Length 6376;
Best Local Similarity 85.7%; Pred No. 0;
Matches 3103; Conservative 0; Mismatches 2; Indels 516; Gaps 5;
1 AAAGTTATCAATTTTACAACTACCGCTATAATCTGCTTGAATCTCTGCAAAAGAGAA 60
866 AAAGTTATCAATTTTACAACTACCGCTATAATCTGCTTGAATCTCTGCAAAAGAGAA 925
61 GACTTTTACCGAGAGAGAGTCTCTGCTGATTCGAAGAACTCAACGAACCAACCCAG 120
926 GACTTTTACCGAGAGAGAGTCTCTGCTGATTCGAAGAACTCAACGAACCAACCCAG 985
121 TTCTCATATATCGTTTAA-----138
986 TTCTCATATATCGTTTAAAGTAAATGATCGGCAATCTGTGTTCTCATTTGTGTGTTT 1045
139 -----138
1046 TGTGTTTGTGTTAGGTTTACAAAGATACCTGAGATTAGTTTGTGTTTGTGTTT 1105
139 --AGGATTACAAAGCTAATCGAGATGAATCCAGAGAAATTCACACACAAAGCAACG 196
1106 TCAGGATTACAAAGCTAATCGAGATGAATCCAGAGAAATTCACACACAAAGCAACG 1165
197 AGCAATTCCTACAGCTCATGAGCTAGCTGTGAATGCGAGGACATGCTCAATTCCTCTT 256
1166 AGCAATTCCTACAGCTCATGAGCTAGCTGTGAATGCGAGGACATGCTCAATTCCTCTT 1225
257 TGCATTTAGCTGGTGTGTTGATCTCTGATCCACCGGTATATTTCTCTCAAGCAATCTCTA 316
1226 TGCATTTAGCTGGTGTGTTGATCTCTGATCCACCGGTATATTTCTCTCAAGCAATCTCTA 1285
317 GTCCGGTGGCAGAGACGAGCTCAATCTGCTGAAAGAGTGTATCAATCAAGCTTGAAGA 376
1286 GTCCGGTGGCAGAGACGAGCTCAATCTGCTGAAAGAGTGTATCAATCAAGCTTGAAGA 1345

```

```

377 AGCTTCTCTTCACAATCTCCTCCACCTGATGATATTCAGCGAGTCTTAGTCTTATTAAGG 436
1346 AGCTTCTCTTCACAATCTCCTCCACCTGATGATATTCAGCGAGTCTTAGTCTTATTAAGG 1405
437 TCATTCTCTGCTGCTCAAGCTGCTCAGAAAGTCAAGAGTCAAGTCTTCAATTTGGCTGTTGAAC 496
1406 TCATTCTCTGCTGCTCAAGCTGCTCAGAAAGTCAAGAGTCAAGTCTTCAATTTGGCTGTTGAAC 1465
497 AGTTGATTATGGTCTCTTCTTGAAGATCTCAATCAGGAGTATTTGTTGAACGAAGTGGTG 556
1466 AGTTGATTATGGTCTCTTCTTGAAGATCTCAATCAGGAGTATTTGTTGAACGAAGTGGTG 1525
557 TAGCGACGGCAGGGTAAAGTCTGAGGTTGAGAGCTTCTGCTGGAAGAGGAGGAGGAGGAG 616
1526 TAGCGACGGCAGGGTAAAGTCTGAGGTTGAGAGCTTCTGCTGGAAGAGGAGGAGGAGGAG 1585
617 TTGAGAGTCTCTCAGGGGACACAAATTTTCAAGCTTTAAAGATTTATGGAAGAGATTTGG 676
1586 TTGAGAGTCTCTCAGGGGACACAAATTTTCAAGCTTTAAAGATTTATGGAAGAGATTTGG 1645
677 TTGAGCAAGCAGGGAAGCTTGATCTCTGCTGATTTGCTGCTGATGAGGAGATTTAGAGATCG 736
1646 TTGAGCAAGCAGGGAAGCTTGATCTCTGCTGATTTGCTGCTGATGAGGAGATTTAGAGATCG 1705
737 TGAGGATTTCTTTCAGGAGAAACGAAGAACTCTGCTGCTGCTTATTTGAGAGGACGAGGATTTG 796
1706 TGAGGATTTCTTTCAGGAGAAACGAAGAACTCTGCTGCTGCTTATTTGAGAGGACGAGGATTTG 1765
797 GTAAACAGCTGCTGTTGAGGTTTATGAGCAAAAGGATTTGGAAGAGATTTGCCCCAACA 856
1766 GTAAACAGCTGCTGTTGAGGTTTATGAGCAAAAGGATTTGGAAGAGATTTGCCCCAACA 1825
857 GTCTTACTGATGAGATTAATTTCTGTTGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 916
1826 GTCTTACTGATGAGATTAATTTCTGTTGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1885
917 ACCGAGGAGATTTGAAAGAGGTTGAAATCTCTTTTGAAGAGGTTGAGGACGCTGGAAG 976
1886 ACCGAGGAGATTTGAAAGAGGTTGAAATCTCTTTTGAAGAGGTTGAGGACGCTGGAAG 1945
977 GCAAAGTATTTCTTTTATTTGATGAGATTCATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1036
1946 GCAAAGTATTTCTTTTATTTGATGAGATTCATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2005
1037 GGTGATGATGAGCTTATCTGTTCAAGCCCATGTTAGCTAGAGGAGGAGCTTCGATGCA 1096
2006 GGTGATGATGAGCTTATCTGTTCAAGCCCATGTTAGCTAGAGGAGGAGCTTCGATGCA 2065
1097 TTGCTGCTACAAAGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTTGA 1156
2066 TTGCTGCTACAAAGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTTGA 2125
1157 GGAAGTTTCAAAGTCTATGTTGCGAGGACCAAGTGTGCTGACACATTTAGTATCTTTA 1216
2126 GGAAGTTTCAAAGTCTATGTTGCGAGGACCAAGTGTGCTGACACATTTAGTATCTTTA 2185
1217 GAGAGCTCAGGAGAGATGAGGAGCATGCTGCTGGAATCCAGACAGAGCTTTA 1276
2186 GAGAGCTCAGGAGAGATGAGGAGCATGCTGCTGGAATCCAGACAGAGCTTTA 2245
1277 TAAATGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311
2246 TAAATGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2305
1312 -----G 1312
2306 GCTGATGTTTATGTTTATGATGATGTTGTTTGTGATGATGATGATGATGATGATGATGATGATGAT 2365
1313 GTGGGCAATTTACCGGATTAAGCAATTTGTTGTTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1372
2366 GTGGGCAATTTACCGGATTAAGCAATTTGTTGTTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2425

```

Db	3506		CAC	TCTGCTCC	AAGTTTTGG	ATGATGGT	CAATTTGA	CAGACGGG	CAAGCAGG	CACAGTCG	CA	3565
Qy	2289		TTT	CAGGAACT	CGGTGAT	AATCAT	GTACATCAAA	ACCCTTGGT	GCTCTGAA	CACCTCCT	TG	2348
Db	3566		TTT	CAGGAACT	CGGTGAT	AATCAT	GTACATCAAA	ACCCTTGGT	GCTCTGAA	CACCTCCT	TG	3625
Qy	2349		GCT	TAATCGG	GAAGTAA	CAATGGA	ATGGCCCGG	GAATGTGT	GAATCGG	GAAGTGA	GA	2408
Db	3626		GCT	TAATCGG	GAAGTAA	CAATGGA	ATGGCCCGG	GAATGTGT	GAATCGG	GAAGTGA	GA	3685
Qy	2409		ACAT	TTCCAG	ACCAGAGCT	CTTTGAA	CAGGCTTGAC	GAGATTTGGT	GTGTTTCA	CCCCCTTC		2468
Db	3686		ACAT	TTCCAG	ACCAGAGCT	CTTTGAA	CAGGCTTGAC	GAGATTTGGT	GTGTTTCA	CCCCCTTC		3745
Qy	2469		ACAT	GACCG	ATTGAGG	AAGTAGCT	CTGGCTTCAAA	TGAAAGAGCTTG	CTCTCCG	CTTGC		2528
Db	3746		ACAT	GACCG	ATTGAGG	AAGTAGCT	CTGGCTTCAAA	TGAAAGAGCTTG	CTCTCCG	CTTGC		3805
Qy	2529		TGA	AGAGG	AGTTGCTTT	TGGCAGTCA	CTGATGCTGCTTT	GGCATATATAT	CTTTGGC	CAGAGAG		2588
Db	3806		TGA	AGAGG	AGTTGCTTT	TGGCAGTCA	CTGATGCTGCTTT	GGCATATATAT	CTTTGGC	CAGAGAG		3865
Qy	2589		TTAT	GA	CCCC-----	-----	-----	-----	-----	-----		2597
Db	3866		TTAT	GACCGG	TAA	GTCTCTT	CCATGATC	GAGATGCATAT	CATTGTTG	ATCTCTC	CAGG	3925
Qy	2598		-----	-----	-----	-----	-----	-----	-----	-----		2604
Db	3926		CTG	GTGTTG	TTCTGTT	GTGACTG	ACTTCTTG	TGGTTCCTT	CTTTGTTG	TGTTAG	GTGTAT	3985
Qy	2605		GGT	GCTAGG	CCTATA	AGGAGAT	GGATGGG	AGAAGGTG	GTGACAGAA	CTGTCA	AAAGATG	2664
Db	3986		GGT	GCTAGG	CCTATA	AGGAGAT	GGATGGG	AGAAGGTG	GTGACAGAA	CTGTCA	AAAGATG	4045
Qy	2665		GTT	GTGGT	GAGG	AAATCGAT	GAAAACTCA	CTGTTTACATAG	TGACGAGG	CGCTGGT	GTAT	2724
Db	4046		GTT	GTGGT	GAGG	AAATCGAT	GAAAACTCA	CTGTTTACATAG	TGACGAGG	CGCTGGT	GTAT	4105
Qy	2725		CTT	GTGTAC	CGGGTG	AGAAAGT	GCGAGGTCTAGT	GGAGCTTCAAC	CAGGCAAGA	AGTCA	GAT	2784
Db	4106		CTT	GTGTAC	CGGGTG	AGAAAGT	GCGAGGTCTAGT	GGAGCTTCAAC	CAGGCAAGA	AGTCA	GAT	4165
Qy	2785		GTG	CTGATT	CATATTT	GTCTAA	CGGGCCAAAG	AGAAGTGA	CTCAGG	CGGTGA	AGAAG	2844
Db	4166		GTG	CTCATT	CATATTT	GTCTAA	CGGGCCAAAG	AGAAGTGA	CTCAGG	CGGTGA	AGAAG	4225
Qy	2845		ATC	GAGATCC	AGG	AAATAG	AAGATGACG	ATAATGAGG	AAATGATCG	AGGAAATAA	AGTCTT	2904
Db	4226		ATC	GAGATCC	AGG	AAATAG	AAGATGACG	ATAATGAGG	AAATGATCG	AGGAAATAA	AGTCTT	4285
Qy	2905		GAC	CTCCAA	CGTGTG	AGTTCTTTT	TGAGGTTGTG	ATGTTGATG	ATGATG	ATGATG	ATGATG	2964
Db	4286		GAC	CTCCAA	CGTGTG	AGTTCTTTT	TGAGGTTGTG	ATGTTGATG	ATGATG	ATGATG	ATGATG	4345
Qy	2965		CTG	TTTTTGG	TTTTT	TGACTCTTTT	TGCTAG	ACTATAAT	GGTGTG	GAACTGTAAA	AAAGTCTAC	3024
Db	4346		CTG	TTTTTGG	TTTTT	TGACTCTTTT	TGCTAG	ACTATAAT	GGTGTG	GAACTGTAAA	AAAGTCTAC	4405
Qy	3025		TTT	TAATATG	TTTTT	CGTATATG	TAAATCA	TTTTTACAT	TAAGAAAT	TCCATG	TACGTC	3084
Db	4406		TTT	TAATATG	TTTTT	CGTATATG	TAAATCA	TTTTTACAT	TAAGAAAT	TCCATG	TACGTC	4465
Qy	3085		ATC	TTTTTGG	GGTTTTT	TTTTCATA						3105
Db	4466		ATC	TTTTTGG	GGTTTTT	TTTTCATA						4486

AC020579 50821 bp DNA linear PLN 19-JAN-2001
Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence,
 complete sequence.

SESSION
AC020579.5 GI:12324896
WORDS
HTG.
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 50821)
Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Roming, C.M., Koo, H., Fujii, C.Y., Uterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence
Unpublished
2 (bases 1 to 50821)
Lin, X. and Kaul, S.
Direct Submission
Submitted (03-JAN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 50821)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280787.
Address all correspondence to: at@tigr.org

BAC clone F1017 is from Arabidopsis thaliana chromosome 1
clone.
The orientation of the sequence is from SP6 to T7 end of the BAC

Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Michaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
GeneSplicer (Michaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers

FEATURES
source

1. 50821
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F1017"
complement(9837..11757)
/gene="F1017.1"
/note="similar to GB:AA049376 from [Glycine max]"
complement(join(<9837..10159,10243..10343,10438..10485,
10576..10693,10800..10995,11127..11248,11404..>11757))
/gene="F1017.1"
complement(join(10052..10159,10243..10343,10438..10485,
10576..10693,10800..10995,11127..11248,11404..>11757))
/gene="F1017.1"
/codon_start=1
/product="putative choline kinase; 11757-10052"
/protein_id="AAG52400.1"
/db_xref="GI:12324898"
/translation="MGTEKNVKNQVRLPREVKALQAIASWEDVIDSKALQVTL
KGAMTNVFIQKWPTRKPGSRKVLVRIYGEVGEIIFDREDEIRTEFFMSKHGGLL

repeat_region
repeat_region
gene

gene

mRNA

CDS

gene

mRNA

CDS

gene

mRNA

CDS

LGRFGNGRIEEFLHARTLSACDLRDEISGRITATRMKEFHGLEMPKAKALLWDLRLN
WLTACKRLASPEAEKFRDLVMEEMENLKESLFDNDENIGCHNDLQYGNIMDEET
KAITIIDIEYSCINPVADIANHFCEMADYHETPHIMDISKYFGEVERQKFLKTYM
SYSDKPSDITMVKLLDEKVEKTYLASHLWGLWGLIISEHVNDISDFYMEYARQFEQY
WLTQRLLAASBHK"
complement(12470..12501)
/rpt_family="AT rich"
complement(13045..13067)
/rpt_family="(A)n"
13093..16423
/gene="F1017.2"
/note="identical to heat shock protein 101 GB:AAF26423
from [Arabidopsis thaliana]"
join(<13093..14242,14327..14508,14649..15274,15359..15836,
15941..16240)
/gene="F1017.2"
/codon_start=1
/product="heat shock protein 101; 13093-16240"
/protein_id="AAG52410.1"
/db_xref="GI:12324908"
/translation="WNPBKFTHKTNETIATAHELAVNAGHAQFTPLHAGALISDPTG
IFPQAISSAGGENAQAQSAERVINQAKLPSPDPDDIPASSSLIKVIRAAQAQKS
RGDTHLVDQLIMGLLEDQIRDLNVEGVATARVSEKVEKRGKGVESASGDTN
FOALATYGRDLVEQAGKLDPVIGRDEEIRRVRIILSRRTKNPVLIGPGVGTAVVE
GLAQRIVGDVPSLTDYRLISLDMGALVAGAKYGEFEERLKSVLKEVEAEKGVIL
FDEIHLVIGAGTSGMDANLFPKPLARGQLRCIGATTILEEYRKYVEKQAEPRFF
QQVYVABSPVDTTISILRLGKEKVEGHGVRIQDRALNAAQSLARYITGRHLPDKAI
DIVDACAQNVQDQPEIDNLRKMQLEIELHALEREKDKASKARLIJEVRKELD
DIRLQKLPITMKYKKEKIDEIRLKKRELMFSLQEAERRYDLARAADRLIGAIO
EVESALQOEGTSSSENVMLTENGVPHIAEVVSWTGTIPVRLGQNKERLIGIADR
LHKRVQGVQANVAYSEALISRAGLGPRPQPTSGSEFLGPTGVGKTELAKALAEOLF
DENLLRDMSEVMEQSVSKLIGAPPEYVGEHEGQQLTEAVRRPVCVILFDEVEK
ACHAVFNTLLQVLDGRLTQDGTVDTPNSVIMTSLNAGHHLAHLGTLGKVTMEVAR
DCVMEVRKHFPPELLNRLDEIVDFPLSHQIRKARLQKMDVAVRAERAGVALAVT
DAALVDILAESYDVPVYGARPTRRMKEKVTLSKMWVREIDENSTVYIDAGAGDLV
YRVESGLVDASTGKSDVLIHLANGPKRSDAAQAVKMKRIEIEDDDNEEMIED"
complement(16481..17587)
/gene="F1017.3"
/note="similar to unknown protein GB:AAD21437 from
[Arabidopsis thaliana]"
complement(join(<16481..17251,17318..>17587))
/gene="F1017.3"
complement(join(16481..17251,17318..17587))
/gene="F1017.3"
/codon_start=1
/product="unknown protein; 17587-16481"
/protein_id="AAG52411.1"
/db_xref="GI:12324909"
/translation="MAIGPSSVIIIVGIIILATVQSILKPPPPNLCSPGPPITAPR
IKRDGRHLAYKGLPREKAKHIVFTHGSDSCKHDAVFATLLSPDLVQGRGVTVMS
FDKPGYGESDPDIPRTKSLALDIELADLSLGSKFYVIGKSMGQGAQGLKTYTH
RLAGVTLVAVVNYRNLPLNISTEGFNQKQDQAVRAVHAHPMLIYWNQTNWF
PGSSVNRDGGVLSQPDKDIILKLGSSSRKPLAEVRQOQIHESINRDMIYVFGHWEFD
PLELENPFLNRGSSVHLWQGDMDLVPTVLTQRYADKLPLWHLHYEVAGGHHFFPLAKG
VYDEIVKTAALNH"
complement(18687..21119)
/gene="F1017.4"
/note="similar to unknown protein GB:AAD21437 from
[Arabidopsis thaliana]"
complement(join(<18687..19284,20351..20547,20799..>21119))
/gene="F1017.4"
complement(join(18687..19284,20351..20547,20799..21119))
/gene="F1017.4"
/codon_start=1
/product="unknown protein; 21119-18687"
/protein_id="AAG52412.1"
/db_xref="GI:12324910"
/translation="NASDSRKSLSKPSRKSFLPPSVVIVIGIIVALTQYKSKPP
QPKLCGSSGPPITAPRIKLQDGRYLAYPEKGLPREKANRIVFIHGSDCRHDVFA
TLLSPDLVEELGVNMFVDFPCYCESDPHSPKTRPSLVSDIEELDDQLSLGSKFVIG

13065 TTTGAGGATTTACAAAGCTTAATCGAAGATGAATCCAGAGAAATTCACACACAGACAAA 13124

13195 CGAGACAATTTGCTACAGCTCATGAGCTAGCTGTGAATGAGACACATGCTCAATTCACCTCC 254

13195 CGAGACAATTTGCTACAGCTCATGAGCTAGCTGTGAATGAGACACATGCTCAATTCACCTCC 13184

255 TTTGCAATTTAGCTGGTGCTTTGATCTCTGATCCACCCGGTATATTTCTCAGACAAATCTC 314

13185 TTTGCAATTTAGCTGGTGCTTTGATCTCTGATCCACCCGGTATATTTCTCAGACAAATCTC 13244

315 TAGTGCCGGTGGCGAGAACGACAGCTCAATCTGCTGAAAGAGTGATCAATCAAGACCTTGAA 374

13245 TAGTGCCGGTGGCGAGAACGACAGCTCAATCTGCTGAAAGAGTGATCAATCAAGACCTTGAA 13304

375 GAAGTTCTCTTCACAATCTCTCTCAACTGATGATATTTCCAGCGAGTTCTAGTCTTTATTAA 434

13305 GAAGTTCTCTTCACAATCTCTCTCAACTGATGATATTTCCAGCGAGTTCTAGTCTTTATTAA 13364

435 GGTCAATTCGTCTGCTCTCAAGCTGCTCAGAAGTCAACAGAGTGATCTCATTTGGCTGTTGA 494

13365 GGTCAATTCGTCTGCTCTCAAGCTGCTCAGAAGTCAACAGAGTGATCTCATTTGGCTGTTGA 13424

495 CCAAGTTGATTAAGGTCCTTTGAGAGATCTCAAAATCAAGGATTTGTTGAAAGAGTCGG 554

13425 CCAAGTTGATTAAGGTCCTTTGAGAGATCTCAAAATCAAGGATTTGTTGAAAGAGTCGG 13484

555 TGTAGCGACGCGAGGGTAAAGTCTCAGGTTGAGAGCTTCGTGGGAAGAGGGAAGAA 614

13485 TGTAGCGACGCGAGGGTAAAGTCTCAGGTTGAGAGCTTCGTGGGAAGAGGGAAGAA 13544

615 AGTTGAGAGTCTTTGAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTT 674

13545 AGTTGAGAGTCTTTGAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTT 13604

675 GGTGAGCAGCAGGAGAGCTTGATCTCTGTGATTTGCTGATGAGGAGATTTAGAGAGT 734

13605 GGTGAGCAGCAGGAGAGCTTGATCTCTGTGATTTGCTGATGAGGAGATTTAGAGAGT 13664

735 CGTGAGGATCTTTTCGAGGAGAACGAAACAATCTCTGCTTATTTGAGAGCCAGGAGT 794

13665 CGTGAGGATCTTTTCGAGGAGAACGAAACAATCTCTGCTTATTTGAGAGCCAGGAGT 13724

795 TGGTAAACAGCTGTGTTGAGGTTTGAACAAGAGATTTGCAAAAGATTTGAAAGGAGATGTCGCCAA 854

13725 TGGTAAACAGCTGTGTTGAGGTTTGAACAAGAGATTTGCAAAAGATTTGAAAGGAGATGTCGCCAA 13784

855 CAGTCTTACTGATGTAGATTAATTTTCGTTGGACATGGTGCGTTAGTTGCTGTGCTTAA 914

13785 CAGTCTTACTGATGTAGATTAATTTTCGTTGGACATGGTGCGTTAGTTGCTGTGCTTAA 13844

915 ATACCGAGGAGGTTTGAAGAAAGGTTGAAATCTGTTTTGAAAGAGTTGAGGACGCTGA 974

13845 ATACCGAGGAGGTTTGAAGAAAGGTTGAAATCTGTTTTGAAAGAGTTGAGGACGCTGA 13904

975 AGGCNAAAGTGATCTCTTTTATTTGATGAGATTCATTTGGTTCTTGGTGCTGGCAAAACTGA 1034

13905 AGGCNAAAGTGATCTCTTTTATTTGATGAGATTCATTTGGTTCTTGGTGCTGGCAAAACTGA 13964

1035 AGGGTCGATGATGATCAGCTAATCTGTTTCAAGCCCATGTTAGCTTAGAGGGCAGCTTCGATG 1094

13965 AGGGTCGATGATGATCAGCTAATCTGTTTCAAGCCCATGTTAGCTTAGAGGGCAGCTTCGATG 14024

1095 CATTTGGTGCTCAACCGTTGAGAAATACAGGAAATATGTTGAGAAAGATGCTGCCCTTTGA 1154

14025 CATTTGGTGCTCAACCGTTGAGAAATACAGGAAATATGTTGAGAAAGATGCTGCCCTTTGA 14084

1155 GAGGAGGTTTCCAAACAAGTCTATGTTGCGGAGCCAAAGTGTGCTTCAGACCACTTAGTATCT 1214

14085 GAGGAGGTTTCCAAACAAGTCTATGTTGCGGAGCCAAAGTGTGCTTCAGACCACTTAGTATCT 14144

1215 TAGAGGACTCAAGGAGAGTATGAGGACATCATGTGTGCGAATCCAAAGACAGAGCTCT 1274

Qy	2120	-----	2119
Db	15285	TATCTTTTAGTTTTCAATATCTCTGATTATGAAATAATTTTTTCTTACTTACACGATTA	15344
Qy	2120	-----GGTATGTTGGTCACGAGGAAGGTGGACAACTAACTCAGAGCTGTGAGG	2166
Db	15345	CTACACTTTCTCAGGTATGTTGGTCACGAGGAAGGTGGACAACTAACTCAGAGCTGTGAGG	15404
Qy	2167	AGGCGACTTATTTGCTCATACCTCTTTGATGAAGTGGAGAAAGGCTCATGTTGCTGTCTTC	2226
Db	15405	AGGCGACTTATTTGCTCATACCTCTTTGATGAAGTGGAGAAAGGCTCATGTTGCTGTCTTC	15464
Qy	2227	AAACACTCTGCTCCAAAGTTTTTGATGATGTCGATTGACAGACGGGCAAGCGCAGGACAGTC	2286
Db	15465	AAACACTCTGCTCCAAAGTTTTTGATGATGTCGATTGACAGACGGGCAAGCGCAGGACAGTC	15524
Qy	2287	GATTTCCAGGAACCTCGGTGATAATCATGACATCAAACTTTGGTCTGAAACACCTCTTGCA	2346
Db	15525	GATTTCCAGGAACCTCGGTGATAATCATGACATCAAACTTTGGTCTGAAACACCTCTTGCA	15584
Qy	2347	GGGCTAACTGGGAAAGTAACATGGAAGTGGCCCGGACTGTGTGATGGGAGGTGAGG	2406
Db	15585	GGGCTAACTGGGAAAGTAACATGGAAGTGGCCCGGACTGTGTGATGGGAGGTGAGG	15644
Qy	2407	AAACACTTTCAGACACAGAGCTCTTTGAAACAGGCTTGACGAGATTTGGTGTTCGACCCCTT	2466
Db	15645	AAACACTTTCAGACACAGAGCTCTTTGAAACAGGCTTGACGAGATTTGGTGTTCGACCCCTT	15704
Qy	2467	TCACTAGACAGTTGAGGAAAGTAGTCGCGCTTCAATGGAAGAAGTTCGCTGTCCGGCTT	2526
Db	15705	TCACTAGACAGTTGAGGAAAGTAGTCGCGCTTCAATGGAAGAAGTTCGCTGTCCGGCTT	15764
Qy	2527	GCTGAAAGAGGAGTTCCTTTGGCAGTCACTGATGCTTCCTTCGACTATATCTTTGCACAG	2586
Db	15765	GCTGAAAGAGGAGTTCCTTTGGCAGTCACTGATGCTTCCTTCGACTATATCTTTGCACAG	15824
Qy	2587	AGTTATGAACC-----	2597
Db	15825	AGTTATGAACCGGTAAAGTCTCTTCACTGATGCAGAAATGCATATCATTTGATCTCTCA	15884
Qy	2598	-----GGTGT 2602	
Db	15885	GGCTGTTGTTCTGTGTACTGACTCTGTGTGTGCTTCTTTGTTTGTGTTAGGTGT	15944
Qy	2603	ATGGTGTAGGCCCTATAAGGAGATGATGCAGAGAAAGGTGTGTGACAGAACTGTCAAAGA	2662
Db	15945	ATGGTGTAGGCCCTATAAGGAGATGATGCAGAGAAAGGTGTGTGACAGAACTGTCAAAGA	16004
Qy	2663	TGTTGTGTGTGAGGAAATCGATGAAAACTCCACTGTTTACATAGATGACGGCGCTGTG	2722
Db	16005	TGTTGTGTGTGAGGAAATCGATGAAAACTCCACTGTTTACATAGATGACGGCGCTGTG	16064
Qy	2723	ATCTGTGTACCGGCTAGAAAAGTGGAGGTCTAGTGACGCTTCAACAGCGAAGAGTCAAG	2782
Db	16065	ATCTGTGTACCGGCTAGAAAAGTGGAGGTCTAGTGACGCTTCAACAGCGAAGAGTCAAG	16124
Qy	2783	ATGTGCTGATTCATATTTGCTAAACGGGCCAAAGAGAAGTGAATCGACTCAGGCGGTGAAGA	2842
Db	16125	ATGTGCTGATTCATATTTGCTAAACGGGCCAAAGAGAAGTGAATCGACTCAGGCGGTGAAGA	16184
Qy	2843	AGATGAGGATCGAGGAAATAGAAAGATGACATTAATGAGGAAATGATCGAGGATTAAGTTC	2902
Db	16185	AGATGAGGATCGAGGAAATAGAAAGATGACATTAATGAGGAAATGATCGAGGATTAAGTTC	16244
Qy	2903	TTGACCTCCAAACGTTGTGAGTCTCTTTTGGGCTGTGTGATTTGATGACTTTGAGCTTTCAATG	2962
Db	16245	TTGACCTCCAAACGTTGTGAGTCTCTTTTGGGCTGTGTGATTTGATGACTTTGAGCTTTCAATG	16304
Qy	2963	CTCTGTTTGGTTTTGACTCTTTTGGTAGACTATATGATGGTGGAACTGTGAAAAAGTCT	3022
Db	16305	CTCTGTTTGGTTTTGACTCTTTTGGTAGACTATATGATGGTGGAACTGTGAAAAAGTCT	16364

3023 ACTTATATATGTTTCGATATGTAATCAATTTACATAAGAAATCCATTGTAGCTCCCA 3082
 16365 ACTTATATATGTTTCGATATGTAATCAATTTACATAAGAAATCCATTGTAGCTCCCA 16424
 3083 TAACTCTTTTGGGGTTTTCATA 3105
 16425 TAACTCTTTTGGGGTTTTCATA 16447

RESULT 6
 SOYSB100HS 3049 bp mRNA linear PLN 15-MAY-1995
 Glycine max heat shock protein (SB100) mRNA, complete cds.
 L35272
 GI:530206
 heat shock protein
 Glycine max (soybean)
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 3049)
 Lee, Y.R., Nagao, R.T. and Key, J.L.
 A soybean 101-kD heat shock protein complements a yeast HSP104
 deletion mutant in acquiring thermotolerance
 Plant Cell 6 (12), 1889-1897 (1994)
 95170290
 7866031
 Original source text: Glycine max cDNA to mRNA.
 Location/Qualifiers
 1..3049
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 1..3049
 /gene="SB100"
 136..2871
 /gene="SB100"
 /note="100 kba"
 /codon_start=1
 /product="heat shock protein"
 /protein_id="AA56338.1"
 /db_xref="GI:530207"

translation="MNPKEFKTHKNEALASAHELAMSSGHAQLPIHLAHLISDPNG
 IFVLAINAGGEEARAVERNLQALKCLPCSPPPDEVPASTNLVRAIRRAQAQK
 SRGTRILVQLILGILEDSQIGDLKEAGVAVAKVESEVDKLRKKGKVEASGDT
 NFOALTYGRDLVBOACKLPVIGRDEEIRRVRIILSRRTKNPNVLVGEVKGKTAVV
 EGAQRIVRGDVPNSLADVRLIADMGALVAGAKYRGEFEERLKVLEVEREAGKVI
 LFIDEHLVLGARTESMDANLFPKPLARQQLRCIGATLSEYKYVEKDAEFRR
 FQVFVAEPPSVDDTISILRGKERYEGHGVRIODRALVMAQLSNRYITGRHLPDKA
 IDLVDEACANVRVQLDSQPEIDNLERKMQLELVHLEKXKASKARLVEVRKEL
 DLRDLQAPLMVKYRKERKRVDEIRLRKKREBELFALQEAERYDLAADIIRYKAI
 QEVETALIQLEGSTEENMLTETVPEQIAEVVSRWTGIPVTRIGQNEKERLIGLGR
 LHSRVQDQAVNAEAVLRSRAGLRPOOPTGSLPLGPTGCTGKTELAKALAEOLF
 DNENLRIIDSMXEMHESVRLIGAPGYVGHGEGQLTEAVRRPYSVWLFDEYK
 AHTSVFNLLQVLDGDTGGGTVDVFNVTIIMTSLNGAHLGLSLGKCTMVAR
 DRWQVRRQRFPELLRLDEIVFPLSHDQLRKVARLQMDKDVASRLAEKGLALAVT
 DAALDYLSSEYDFYKARIRWLEKKVYTELRLVREEIDENSTVTVIDAPNGGE
 LVYRVKNGGVNPTTQCKSDILIQIPNGHAPKTDVAQVKKMKIEIDDEEMEE"
 850 a 564 c 863 g 772 t

Query Match 52.8%; Score 1638.4; DB 8; Length 3049;
 Best Local Similarity 75.9%; Pred. No. 0;
 Matches 2094; Conservative 0; Mismatches 646; Indels 18; Gaps 5;
 161 AGATGAATCCAGAGAAATTCACACACAGACAAACGAGACAATTTGCTACAGCTCATGAGC 220
 134 AAATGAATCCTGAGAAGTTTACTCACAGACTAATGAAGCTCTTGGCTCGGCCACGAGC 193
 221 TAGCTGTAATGCAGGACATGCTCAATTCCTCTTGCATTTAGCTGGTGGCTTTGATCT 280

194 TCOCGATGATGTTCAGCCACACGCGCAATTCCTCCATCCATTCCTGCCCCACGCGCTGATTT 253
 281 CTGATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTGC---CGTGGCGAGAACGAG 337
 254 CCGATCCCAACCGGTATATTTCTGCTGCTAGCATTAACAGCGCGGCGCGCGAGGAATCG 313
 338 CTCAATCTGCTCAAAAGAGTGAATCAATCAAGCCCTTGAAGAAGCTTCTTTTCAAAATCTCTC 397
 314 CACGCGCGCTGAGCGAGTGTGTAACACAGGCTCTGAAGAAGCTACCTGCCAGTCCCCTC 373
 398 CACCTGATGATATTCAGCGAGTCTTAGTCTTATTAAGGTCAATTCGTCTGCTCAAGCTG 457
 374 CCGCGGACGAGTGCAGCGAGCAACCACTCTGAGAGGCCATCAGGAGACACAGCGCG 433
 458 CTCAAGATCACAGAGTACTACTATTTGGCTGTGACAGTTGATTAATGCTCTTCTTG 517
 434 CCGAAAATCACGTGGCGACACGCGTTTGGCCGTTGATCAGTTGATCTCGAATCCTCG 493
 518 AAGATCTCAAAATCAGGATTTGTTGAACGAGTCTGTTAGCGACGCGAGGGTAAAGT 577
 494 AAGATCTCCAAATCGAGACCTGTGTAAGAGGAGCGGGGTTGCGGTGCGAAGGTAGAAT 553
 578 CTGAGTTGAGAGCTCTGCGGAGAGGAGGAGAAAGTTGAGAGTCTTCAGGGACA 637
 554 CGAAGTGTGATTAAGCTTCTGGAAGAGGAGGAGGAGGTTGAGAGCGCTTCCCGGGATA 613
 638 CAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGAGCAACGAGGAGGAGCTTG 697
 614 CGAATTTCAAGCTTTGAGACTTATGGCGTGACCTTGTGTAACACGCGGAGGAGCTCG 673
 698 ATCTCTGATTTGCTGATGAGGAGATTAAGAAGTCTGAGGAGTCTTTCGAGGAGAA 757
 674 ACCCTGTTTATTTGCGCGCTGACGAGAGATTAAGAAGGTTGAGAGGATTTATCAGGAGA 733
 758 CGAAGAACAACTCTCTGCTTATTGAGAGGACGAGGTTGTTAAACACGCTCTGTTGAAG 817
 734 CTAGAAACAAACCCGGTCTCTGTTGGAGAACCGGTTGGGAAAACTCGGTTTGGGAAG 793
 818 GTTTAGCACAAGAGATTTGTAAGGAGATGTGCCCAACAGCTTTACTGATGTGAGATTAA 877
 794 GGTGGCACAAGAGATTAAGAGGGGATGTTCCAAGCAACCTTCTGATGTGAGGCTTA 853
 878 TTTGCTTGGACATGCGTCTGTTAGTTGCTGCTGCTAAATACCGAGGAGAGTTTGAAGAA 937
 854 TTGCGTTGATATGCGGGCGTTTGTGTCGCGGCTGCAAGTATAGAGCGGAGTTTGAAGAGC 913
 938 GGTTCAAATCTGTTTTGAAAGAGTTGAGGAGCGCTGAAGGCAAGTGAATGTTCTTTATTG 997
 914 GGTAAAGCTGTTTTGAAAGAGTTGAGGAGGCTGAGGAGAGGTGATCTCTTCTATTG 973
 998 ATGAGATTCATTTGTTCTTGTGCTGCTCAAAACTGAAGGTCGATGATGATCAAGTCAATC 1057
 974 ATGAGATTCATTTGTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1033
 1058 TCTTCAAGCCCATGTTAGCTAGAGGAGCTTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1177
 1034 TATTCAACCTTATGCTTCTCGCGCCAGCTTTAGTGATTTGGTGCCACCAAGCTTTGAGG 1093
 1118 AATACAGGAATATCTTGAAGAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGTCTATG 1177
 1094 AGTACAGAGATTAAGTGGAGAGGATGCTGATTCGAGAGGAGGTTCCACAGGTTTGTG 1153
 1178 TTGCGGAGCCAGTGTGCTGACACCATTAATGATCTTCTTGAAGGACTCAAGGAGATGATG 1237
 1154 TGGCGGAACCTAGTGTGCTGATACCATTAATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1213
 1238 AGGGACATCATGTTGCGAATCCAGGACAGCTCTTATAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1297
 1214 AAGGCCATCAGGTTGTAATTCAGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1273
 1298 CTGCTTACATAACTCGTCCGCTATTTACCGGATAAGCAATTTGATTTGGTGTGATGAGGCTT 1357
 1274 ACCGCTATTAACCTGGGCTCATCTTCTGACAGGCAATTTGATTTGGTGTGATGAGGCTT 1333

1358 GTGCGAATCTGAGACTCCAGCTTGATAGTCAACTGAAGAGATTGATAACCTTGAAGGA 1417

1334 GTGCAATGTTAGGGTTCAACTTGATAGTCAGCTCAGGAAATTGATACTTGAAGGA 1393

1418 AGAGGATGCAGCTGGAATTTGAACTTCACGCTTGGAAAGGAGAAGTAAAGCCAGCA 1477

1394 AGAGAATGCAGCTAGAAGTGAACTTCATGCTCTGGAGAAAGAGAAAGCAAGGCTAGCA 1453

1478 AAGCTCGACTTATAGAGTGGGAAAGAGCTTGATGACCTGAGAGACAAGCTTCAGGCTC 1537

1454 AAGCCCGTCTGTTGAAGTGGGAAAGAACTTGATGACTTGAAGGACAAGCTTCAGCGTT 1513

1538 TCAGATGAATATACAGAAAGCAGAAAGAGAGAAATCGATGAGATTCGAAAGCTTAAACAGA 1597

1514 TGAATAGAGTACCGAAAGAGAAAGAGAGGTTGATGAGATTCGAAGGCTCAAGAGA 1573

1598 AAAGAGAAGAGCTCATGTGTTTCTTTCAGGAGGCAGAAACGAAGATATGACCTTGCAGAG 1657

1574 AAAGAGAAGAGCTCTTTTTCGCTTTTCAAGAGGCTGAGAGAGATATGATTCGCTAGAG 1633

1658 CTGCTGATCTAAGATATGGCGCAATTCAGAAATGGAATCTCGAATTCGCCAACTTGAAG 1717

1634 CTCGACAGCTCGGATATGGAGCAATTCAGAGGTGAAATCTCAATACAACTTGAAG 1693

1718 GAACCTTCTTCGAAAGAAATGATGCTCACAGAAACGTTGGGCTGAAACATCTGCTG 1777

1694 GGA---GCATCGAAGAGAACTGATGTTGACTGAAACTGTTGGAACGAGCAATAGCTG 1750

1778 AGGTTGTGAGCGCTTGACACAGGATTCAGTGCAGAGACTTGGCCAAAATCAGAAGAGA 1837

1751 AGGTTGTGAGCGCTGGACCGGTATACCACTTACAAGGCTTGGCCAAATGAAAGAGAA 1810

1838 GATTGATTTGGTCTTGCTGATAGTTGCTAAGCGGTTGTGGGACAGAAATCAACGGGTAA 1897

1811 GATTGATTTGGAATCTGGTGACAGATTCACACAGCAGAGTTGTAGGACACAGCAAGCTGTA 1870

1898 ATGCAGTTTCTGAGGCAATCTTAAGGTCAAGGCGCAGACTCTGTAGGACCAACACAGCAA 1957

1871 ATGCTGTTGCTGAGGCTGTTGAGATCTAGACTGGGCTAGGAAGACCTCAACACAA 1930

1958 CTGGATCATTTCTATTCTTTGGACCAACTGGTGTGGCAAAACTCAGCTCGCCAAAGGCTC 2017

1931 CTGGTTTCATTCTTATCTCTTGGTCCAACTGGTGTGTTAAACTCAGCTTGCAGAGGCTC 1990

2018 TTGCTGAGCAGCTGTTTGATGATGAAAACTCTTAGTTTCGATTCGATTCGGAATATGCGAAATATA 2077

1991 TTGCTGAGCACTATTTGATAATGAAAAACAGTTGGTGAGAAATTCGATATGTCGAAATACA 2050

2078 TGGAAACAACCTGCTCTCGCTCATTTGGGGCACACAGGGGTATGCTGGTCAACGAGG 2137

2051 TGGAGCAGCACTCGGTTTCAAGATTGATTGGTGACCAACCGGTTATGTTGACATGAGG 2110

2138 AAGTTGGACAACTTAATCTGAGGCTGTGAGAGGCGACCTTATTTGTTGCATCTCTTTGATG 2197

2111 AAGTTGGACAACTTAATCTGAAGCGGTAAAGCGAAGACCTTACAGTGTGGTGTCTTTTGATG 2170

2198 AAGTTGGAAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAGTTTTCGATGATGCTC 2257

2171 AAGTTGGAAGGCAATACATCTGTGTTCAATACCTCTTCTTCAGTTTTCGACATGGGA 2230

2258 GATTGACAGCGGCAAGGCAGACAGTCGATTTCAAGAACTCGGTGATATCATATGACAT 2317

2231 GGTTAATCTGATGGACAAGGCCGTACTGTGGACTTTAGAAACACTGTAATTAATCATGACCT 2290

2318 CAAACCTTTGGTGTGAACACCTCTTTCAGGGCTAACTGGGAAAGTAACATGGAAGTGG 2377

2291 CAAATCTTGGAGCAGACCACTCTCAGTGGACTTTCAGGAAAAATGACCATGCAAGTTG 2350

2378 CCGGGACTGTGTGATTCGGGAGGTGAGGAAACACTTCAGACAGAGCTCTTGHAAAGGC 2437

2351 CTCGGGATTCGGGTAAATCGAGGAGGTGAGAAGGCAAATTTAGGCCAAGTAATGCTGAATCGAC 2410

QY	2438	TTGACGAGATTGGTGTTCGACCCCTTTTCACATGACCAGTTGAGAAAGTAGTCGGC	2497
Db	2411	TTGATGAAATTTGGTATTTTGACCCCTTTTCATGACCAATTTGAGGAAGTTGCAAGGC	2470
QY	2498	TTCAAATGAAAGACGTTGCTCGCGCTTCTGAAAGAGGAGTTGCTTTGGCAGTCACTG	2557
Db	2471	TACAAATGAAGGATGTAGCTAGTTCGTCTTGCGAAGAGGAAATTCCTTGGCAGTCACTG	2530
QY	2558	ATGCTGCTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGGTGTATGFGTGCCTAGCCCTA	2617
Db	2531	ATGCAGCACCTAGACTATATATCTTTCCGAGAGCTATGATCCGGTGTATGGTGTCTAGACCAA	2590
QY	2618	TAAGGACATGATGAGAGAAAGGTGGTGACAGAACTGTCAAGAGATGTTGTTCGCTGAGG	2677
Db	2591	TAGGAGCTGGCTTGAGAGAGAGTGTGACAGATTGCTAGAAATGCTTTGTAAGAGAGG	2650
QY	2678	AAATCGATGAAACTCCACTGTTTACATAGATCAGG-----CGCTGGTGTATCTTTGTGT	2731
Db	2651	AGATTGATGAGAAATCAACGGTTTACATAGATCTGGACCCCAATGGGGCGAGTTGTGTCT	2710
QY	2732	ACCGGTGAGAA---AGTGGAGTCTAGTGCAGCGCTTCAACAGGCAAGATGCAGATGTGC	2788
Db	2711	ACCGTGTAGAAAAGAAATGAGAGGGTGTGTTAATCTCAACTGGGCAAGATCCGATATCT	2770
QY	2789	TGATTCAATTTGCTAAACGGGC---CAAAGAGAGTGTATGCAGCTCAGGCGGTCAAGAGA	2845
Db	2771	TGATTCAGATACCTAATGACATGCACCTAAACTGTCTGTTCAAGCAGTCAAGAAAG	2830
QY	2846	TGAGGATCGAGGAAATAGAAAGATGACGATTAATGAGGAAATGATCGAGGATTAAGTCT	2903
Db	2831	TGAAGATTGAGGAAATGATGATGATGAAATGGAAGATGAATTTTATGATAAAATGT	2888
RESULT 7	AF083343		
LOCUS	Nicotiana tabacum heat shock protein 101 kDa (HSP101) mRNA, complete cds.	3052 bp	linear
DEFINITION			
ACCESSION	AF083343		
VERSION	AF083343.2	GI:11561805	
KEYWORDS			
SOURCE	Nicotiana tabacum (common tobacco)		
ORGANISM	Nicotiana tabacum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamids; Solanales; Solanaceae; Nicotiana.		
AUTHORS	1. (bases 1 to 3052)		
TITLE	Wells, D.R., Tanguay, R.L., Le H. and Gallie, D.R.		
JOURNAL	HSP101 functions as a specific translational regulatory protein whose activity is regulated by nutrient status		
MEDLINE	Genes Dev. 12 (20), 3236-3251 (1998)		
PUBMED	99003076		
REFERENCE	2. (bases 1 to 3052)		
AUTHORS	Gallie, D.R., Tanguay, R.L. and Wells, D.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-AUG-1998) Biochemistry, University of California-Riverside, Riverside, CA 92521-0129, USA		
REFERENCE	3. (bases 1 to 3052)		
AUTHORS	Gallie, D.R., Tanguay, R.L. and Wells, D.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-DEC-2000) Biochemistry, University of California-Riverside, Riverside, CA 92521-0129, USA		
REMARK	Sequence update by submitter		
COMMENT	On Dec 6, 2000 this sequence version replaced gi:3982830.		
FEATURES	<p>Location/Qualifiers</p> <p>1..3052</p> <p>/organism="Nicotiana tabacum"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:4097"</p> <p>1..3052</p> <p>/gene="HSP101"</p> <p>101..2830</p>		
gene			
CDS			

```

/gene="HSP101"
/codon_start=1
/product="101 kDa heat shock protein"
/db_xref="GI:11561806"
/translaton="MNPKEKTHKNEALAGALELALAGAHQAQTPPLHMAVALISDNG
SRGSHLAVQDLILGLLEDSQIGLLKEAGVSASRVKSEVEKLRGKGRKVESASGQK
TFQALNTYGRDLVPAQAKLPVIGRDEIRRVVIRLIRSRKNNPVLIGEPGVGKTAIV
EGLAQRVIRGDPNLADELIALDMGALVAGKRGEPERLKAFLVKEVEAEKQVI
LFIDEHLVLGAGTSGMDAANLFKPLMARGOLRCIGATLLEVKYVEKDAFERR
FOQYVEPSVADPISILGKERYEGHGVKIQDRALVVAQLSRYITGRHLPKA
IDLVEACANVRQDSQDELELLEKRIQLELLEHLEKDKKAKLIEVRKEL
DLEKLOPLNMRKYKKEKRIEDELQKQKDELIALQEAERYDIALRAADLRYGAI
QEVETANLESTAESMTLTPDQPIAEVWSRMTGIPIVSRIGQNEKEKLIQGLGR
LHQVWGQDAVAVASVLRSLRGLPGPOPTGSLFLGPTGVKGTETELAKALABOLK
DDKLMIRIDMSVEYMSVSLRGLPAGPVGVHDEGQLEAVRRPYSVVLDFEVEK
AHPAVFTLLQVLDGRLTGQGRVTPNTSVIIMTSLGAEVLLSLMKKCTMETAR
EMVQEVKQPKPELNRLEIVVDFELSHKQKQVCRVQKMDVALKALERGIALGV
EAPDVLITSESDPYGARPFLRWLERKLVTELSKMLVKEEDENSTVYIDAGVSGKD
LTYREKNGGLVNAATQCKSDILQLPNGPRSDAVQVKQRIEIEIDDEMED"
LSE COUNT      898 a 536 c 836 g 782 t
IGIN

```

```

Query Match      50.4%; Score 1564.4; DB 8; Length 3052;
Best Local Similarity 74.4%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 686; Indels 18; Gaps 5;

147 CAAAGCTAAATCGAAGATGAATCAGAGAAATTCACACACAGACAAACAGACAAATTGC 206
85 CTATAGGTGACTTATTGATCTGGAATAATTCACCAACAGACTAACGAGGCCCTTGC 144
207 TACAGCTCATGAGCTAGCTGTAATGAGGACATGCTCAATTCATCTCTTGTGCAATTGAC 266
145 TGGGCACTCGAGCTAGCACTATCCGACAGGCGATGCTCAATTTACGCTCTGCATATGGC 204
267 TGGTCTTGTATCTGATCCACCGGTATATTTCTTCAAGCAATCTCTAGTGCAGGTG 326
205 TGTGGCTTAATATCTGATCAATGATGTTATTTTCGACAGCGATTTGCAATGCTGGTGG 264
327 ---CGAGAACGAGCTCAATCTGCTGAAAGAGTGATCAATCAAGCCCTTGAAGAGCTTCC 383
265 GAATGAAGAAAGTAGCTAATTCAGTGGAGCGGTATTGAACTCAAGCATGAAGAGCTACC 324
384 TTCAATATCTCTCCACCTGATGATATTCAGCGAGTTCAGTCTTATTAAGGTCAATTCG 443
325 TTCTCAACACCGGCTCTGACGAAATCCCACTAGCACTTCCCTTATCAAGGTGTTACG 384
444 TCGTGTCTCAAGCTCTCAGAGTCCAGAGGTGATCACTATTGGCTGCTTGCACCACTGAT 503
385 CCGAGCAATCTGCGCAGAGTCTCGTGGTGAAGCCATTTAGCAGTGATCAGTTGAT 444
504 TATGGGTCTTCTTGAAGATCTCAAAATCAGGGATTTGTTGAACGAAATCGGTGTAGCGAC 563
445 TTTAGGACTGTGAGAGACTCCCAATTTGAGAGTCTTTTGAAGAAAGCTGGAGTGC 504
564 GCGAGGGTAAAGTCTGAGGTGAGAGCTTCGTGGGAAAGAGGGAAGAAAGTTCGAAAG 623
505 ATCAAGAGTGAATCAGAGGTAGAGAAATTTAGAGGAAAGAGGAAAGAAAGTTCGAAAG 564
624 TGGTTCAGGGGACACAAATTTCAAGCTTTAAAGCACTTATGGAAGAGATTTGGTTCAGCA 683
565 TGGTTCAGGGGACACCAATTTCAAGCACTTCAACACTTATGGCCGTGATCTTTGGGAACA 624
684 AGCAGGGAAGCTTATCTGTGATTTGGTGTGATGAGGAGATTAGAGAGTCTGTGAGGAT 743
625 AGCAGGAAGCTTATCTGTGATTTGGTGTGATGAGGAGATTAGAGAGTCTGTGAGGAT 684
744 TCTTTCAGGAGAGAGGAGAAACAAATCTGTGCTTATTTGGAGAGCCAGAGATTGGTAAAC 803
685 TTTATCAAGGAGGAGACTAAGAACAAACCGGTTCTTATTTGGAGAGCCCGGTGTGGGTAAC 744
804 AGCTGTGCTTGAAGTTTAGCACAAAGGATTTGTAAGGAGATGTGCCCAACAGCTCTTAC 863

```

```

Db      745 AGCAGTGTGTGAAGGCTAGCACAGAGGATTTACGTGCTGATGTTCAAGTAATTTAGC 804
Qy      864 TGATGTGAGATTAAATTTGTTGGGACATCGGTGCGTACGTGCTGCTGCTTAATACCGAGG 923
Db      805 TGATGTTAGGCTTTATAGCATTTGATATGGAGCGGTAGTTGCTGGAGCTTAAGTACAGAGG 864
Qy      924 AGAGTTTGAAGAAAGGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGAGGCGAAAGT 983
Db      865 TGAATTTGAAGAGAGGCTGAAGGCTGTCTGTAAGAAAGTTGAAGAAAGCGGAAAGT 924
Qy      984 GATTTCTTTTATTTGATGAGATTCATTTGTTTCTTGGTGTGCTGCAAAAACCTGAAGGTCAT 1043
Db      925 AATACTTTTCAATTTGACGAGATACATTTATGTTCTGCTGCTGCTGCGACAGAGGCTCTAT 984
Qy      1044 CGATGCACTTAATCTGTTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCCATGCTATTTGGTGC 1103
Db      985 GATGCTGCTTAATCTGTTTAAAGCAATGCTAGCAGAGGCTCAATTTACGGTGCATTTGGTGC 1044
Qy      1104 TACAAAGCTTGAAGAAATACAGGAAATATGTTTGAAGAAAGATGCTGCTTTGAGAGGAGT 1163
Db      1045 AACTACACTCGAGGAGTACAGGAGTATGTTTGAGAAAGGATGCTGCAATTTGAGAGGCGTTT 1104
Qy      1164 CCAACAAAGTCTATGTTGCGGAGCCAAAGTGTGCTGCAACCAATTAGTATCTTTAGAGACT 1223
Db      1105 CCAGCAGGTTGATGTTGCTGAGCCCTAGTGTGCTGACACTATTAGTATTTCTCCGTTGGTT 1164
Qy      1224 CAAGAGAGATGATGAGGACATCATGTTGTGCAATCCAAGACAGAGCTCTTATAAATGTC 1283
Db      1165 GAAGAGAGTATGAAGGCACTATGTTGTCAAATTCAGGACAGAGCTCTTGTAGTGC 1224
Qy      1284 TGCTCAGCTGTCTGCTGTTTACATAAATGTTGCGGCAATTTACCGGATAAAGCAATGATTT 1343
Db      1225 TGCCAGCTCTCATCTCGGTATCATTTACAGGTGCAATCTGCCAGATAGGCTATTGACCT 1284
Qy      1344 GGTGATGAGGCTTCTGCGAATGTGAGAGTCCAGCTGATAGTCAACTCAAGCTGAAGAGATTGA 1403
Db      1285 AGTGTATGAGGCTTGTGCAATGTTAGAGTTGAGTTGATAGTCACTGAGGAAATTTGA 1344
Qy      1404 TAACTTTGAAGAAAGAGATGACGCTGGAATTTGAACTTCAAGCTTTGGAAGAGGAGAA 1463
Db      1345 CATCTTGAAGAGAGAGATTCAGCTAGAGGTTGAACTTCAAGCTCTCGAGAGGAGAA 1404
Qy      1464 GATTAAGCAGCAAAAGCTCGACTTATAGAGGTGCGGAAAGAGCTTGAATGACCTGAGAGA 1523
Db      1405 AGACAAAGCTAGCAAAAGCACGCTCTCATAGAAGTGAAGGAAAGAACTTGTATTTGAGAGA 1464
Qy      1524 CAAGCTTCAAGCTCTCACATGAAATACAGAAAGGAGAAAGAGAGAAATTTGATGAGATTGC 1583
Db      1465 CAATCTCCAACTTTGATGATGAGGTACAGAAAGAGAGAGAGAGATAGATGAGCTGCG 1524
Qy      1584 AAGGCTTAAACAGAAAGAGAGAGCTCATGTTTCTTTTTCAGGAGGCGAGAACCGAAGATA 1643
Db      1525 CAGGCTCAAGCAAAAGCGCATGAGCTCATCTATGCTTTTCAAGAAAGCTGAAAGAGAGATA 1584
Qy      1644 TCACTTTCAGAGAGCTGCTGATCTAAGATATGCGCAATTTCAAGAGTGGAAATCTGCAAT 1703
Db      1585 TGATCTGGCGAGGCGAGAGATCTGAGATNTGGGCAATTTCAAGAGTGGAAATCTGCAAT 1644
Qy      1704 TGCCCAAATTTGAAGCAACTTCTTCTGGAAGAGATGATGATGCTCAAGAAAACGTTGGGCC 1763
Db      1645 AGCAATCTTGAAGATACCTCAGCTGAA ---AGTACATGCTTAACAGAGACTGTGGGTCC 1701
Qy      1764 TGAACACATTTGCTGAGGTGTGAGCCGTTGACAGGGATTCAGTGAAGAGACTGTGGCCA 1823
Db      1702 TGATCAGATTGCGGAAGTTGTGAGTGTGAGCTGATTTCCGGTCTCAAGGCTTGGTCA 1761
Qy      1824 AATGAGAGAGAGAGGTTGATTTGCTTGTGATGAGTTGCTAAGCGGGTTGTGGGACA 1883
Db      1762 GAATGAGAAAGAGAAACTGATTTGGTCTTTGGCGATAGATTGCAACAAAGAGTGTGGCGCA 1821
Qy      1884 GAATCAAGCGGTAAATGCAATTTCTGAGGCAATTTCAAGGTCAAGGGCAGGACTTGGTAG 1943

```

1822 AGATCATGAGTGTAGAGCTGTTGCTGAAGCCGCTGTTAAGGTCACAGAGCTGGTTAGGAAG 1881
 1944 GCACAAACCACTGATCATCTTATTCCTGACCACTGCTGTTGCAAACTGA 2003
 1882 GCCACGACCACTGCTGTTCACTTCTTCTGGGCGCACTGCTGTTGGAAGACAGA 1941
 2004 GTCGCAAGGCTCTGCTGAGCAGCTGTTGATGATGAACACCTCTTAGTTCCGATTGA 2063
 1942 GCTCGCTAAAGCTCTGACAGCAGCTCTTTCATGATGATGAACATGATGATGATGA 2001
 2064 TATGTCGATATATATGACCACTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2123
 2002 CATGTCGAGTATGACCACTCTGTTTCCGCGCTGATGCTGCTGCTGCTGCTGCTGCT 2061
 2124 TCTGCTGACAGGAGGTTGACCACTAATGCTGAGCTGCTGAGGCGGCACTTATTGCT 2183
 2062 TCTGCGCATGATGAGGAGGACCACTTACTGAGCTGTTAGGAGCGGCTTACGATG 2121
 2184 CATATCTTTGATGAGTGGAGAGGCTCATGCTGCTGCTTCAACACTCTGCTCCAAGT 2243
 2122 TGTGCTATTTGATGAGTGGAGAGGCTCATGCTGCTGCTTCAACACTCTGCTCCAAGT 2181
 2244 TTTGATGATGCTGATGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2303
 2182 TCTGATGATGAGAGGTTAAACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2241
 2304 GATAATCATGACATCAAACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2363
 2242 GATTATGATGCTCAAACTTGGAGCAGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTG 2301
 2364 AACATGAGTGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2423
 2302 TACCATGAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2361
 2424 GCTCTGACAGGCTGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2483
 2362 GCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2421
 2484 GAAAGTAGCTCGGCTTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2543
 2422 GCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2481
 2544 TTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2603
 2482 ATTGGCGCTTACTGAGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2541
 2604 TGGTCTAGGCTTAAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2663
 2542 TGGTCAAGACCTATTAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2601
 2664 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2717
 2602 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2661
 2718 TGGTATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2774
 2662 GAAAGATCTAACTTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2721
 2775 GAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2834
 2722 AAAATCTGATATATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2778
 2835 GGTGAGAGATGAGGATCGAGGAAATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2882
 2779 AGTCAAGAAGATGAGGATTTGAAGAAATTTGAAGATGAGATGAGATGAGATGAGATGAG 2826

ULT 8
 62596
 US
 INITIATION
 mRNA, partial cds.
 AY062596

AY062596.1 GI:17065039
 FLI CDNA
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Arabidopsis thaliana
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 1570)
 Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M.,
 Palm, C.J., Bowser, L., Jones, T., Bath, J., Carninci, P., Chen, H.,
 Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
 Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (14-NOV-2001) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 e-mail for correspondence: arab@sequence.stanford.edu
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Nguyen, M.,
 Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J.,
 Bowser, L., Jones, T., Bath, J., Chen, H., Cheuk, R., Chung, M.K.,
 Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P.,
 Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.
 Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed
 equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W.
 (SSP/Stanford) contributed equally to this work as PIs.
 Location/Qualifiers
 1..1570
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="RAFL08-19-P05"
 /notes="This clone is in pBluescript"
 1..1570
 /gene="At1g74310; F1017.2"
 /gene="At1g74310; F1017.2"
 /codon_start=1
 /product="heat shock protein 101"
 /protein_id="AAL32674.1"
 /db_xref="GI:17065040"
 /translation="MNPEKFTHTKNTTATTAHLAVNAGHAQFTPLHLAALISDPTG
 IFPQALSSAGGNAASERVINOALKLPSPDPDDIPASSLLIKVIRRAQAOKS
 RGDTHLAVDQLIMGLLEDSDRLNEVGATVATRVSEKRLGKRGKRVESASGDTN
 FOALTKTGDLVEQAKLDPIVGRDEIRRVVRLISRTKNNPLVIGSPGVGTAVVE
 GLAQRIVKGDVPSNTDVLRLISLDMALVAGAKYRFEERLKSVLKVEVDAEGKVL
 FIDEHLVLGAKTSGSDAANLFPMLARGOLRCIGATTLREYKRYEKDAAPERF
 QQVYAPRSPVPTISILRLGKKEGHEGVRIQDRALNAAQLSARYITGRHLPKAI
 DIVDEACANRVQLDSQPEEDINLERKMQLELHALEREKDKASKARLLIEVRKELD
 DLRLDKLQPTT"
 BASE COUNT 467 a 280 c 407 g 416 t
 ORIGIN
 Query Match 49.7%; Score 1542.4; DB 8; Length 1570;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAGTTATCAATTTTACAACTTACCGCTATATCTGCTTGTGATTTCTCTGCAAAAAGAGAA 60
 Db 27 AAAGTTATCAATTTTACAACTTACCGCTATATCTGCTTGTGATTTCTCTGCAAAAAGAGAA 86
 QY 61 GACTTTTACCGAGAGAAAGTCTCTGCTGCTATTGAGAAACTCAACGAAACCAACCCAG 120

1167	DB	GATGCTGCCTTTGAGAGGAGGTTTCAAACAAGCTTATGTTGCGGAGCGAAGTGTCGCTGAC	12226
1201	QY	ACCATTTAGTATCCTTAGAGGACTTCAAGAGAGAGTATGAGGGACATCATGTGTGCGGAATC	1260
1227	DB	ACCATTTAGTATCCTTAGAGGACTTCAAGAGAGAGTATGAGGGACATCATGTGTGCGGAATC	12886
1261	QY	CAAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTGTTACATAACTGTCGGGCAT	1320
1287	DB	CAAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTGTTACATAACTGTCGGGCAT	1346
1321	QY	TTACCGGATAAAGCAATTGATTTTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTT	1380
1347	DB	TTACCGGATAAAGCAATTGATTTTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTT	1406
1381	QY	GATAGTCAACCTCAAGAGATTGTATAACCTTTAAAGGAAGAGGATGCAAGCTTGGAAAATTGAA	1440
1407	DB	GATAGTCAACCTCAAGAGATTGTATAACCTTTAAAGGAAGAGGATGCAAGCTTGGAAAATTGAA	1466
1441	QY	CTTCAGGCCTTGAAAGGGAGAGGATAAAGCCAGCAAGCTCGACTTATAGAGGTGCGG	1500
1467	DB	CTTCAGGCCTTGAAAGGGAGAGGATAAAGCCAGCAAGCTCGACTTATAGAGGTGCGG	1526
1501	QY	AAAGAGCTTGATGACCTGAGAGACAAGCTTTCAGCCTCTCACGAT	1544
1527	DB	AAAGAGCTTGATGACCTGAGAGACAAGCTTTCAGCCTCTCACGAT	1570

AF133840	LOCUS	AF133840	3084 bp	mRNA	linear	PLN 21-NOV-2001
DEFINITION	Zea mays	heat shock protein HSP101	(HSP101)	mRNA	complete cds.	

ACCESSION AF133840

KEYWORDS .

ORGANISM	Zea mays	Pinus strobus
1. <i>Pinus strobus</i>		
2. <i>Zea mays</i>		
3. <i>Pinus strobus</i>		
4. <i>Zea mays</i>		
5. <i>Pinus strobus</i>		
6. <i>Zea mays</i>		
7. <i>Pinus strobus</i>		
8. <i>Zea mays</i>		
9. <i>Pinus strobus</i>		
10. <i>Zea mays</i>		
11. <i>Pinus strobus</i>		
12. <i>Zea mays</i>		
13. <i>Pinus strobus</i>		
14. <i>Zea mays</i>		
15. <i>Pinus strobus</i>		
16. <i>Zea mays</i>		
17. <i>Pinus strobus</i>		
18. <i>Zea mays</i>		
19. <i>Pinus strobus</i>		
20. <i>Zea mays</i>		
21. <i>Pinus strobus</i>		
22. <i>Zea mays</i>		
23. <i>Pinus strobus</i>		
24. <i>Zea mays</i>		
25. <i>Pinus strobus</i>		
26. <i>Zea mays</i>		
27. <i>Pinus strobus</i>		
28. <i>Zea mays</i>		
29. <i>Pinus strobus</i>		
30. <i>Zea mays</i>		
31. <i>Pinus strobus</i>		
32. <i>Zea mays</i>		
33. <i>Pinus strobus</i>		
34. <i>Zea mays</i>		
35. <i>Pinus strobus</i>		
36. <i>Zea mays</i>		
37. <i>Pinus strobus</i>		
38. <i>Zea mays</i>		
39. <i>Pinus strobus</i>		
40. <i>Zea mays</i>		
41. <i>Pinus strobus</i>		
42. <i>Zea mays</i>		
43. <i>Pinus strobus</i>		
44. <i>Zea mays</i>		
45. <i>Pinus strobus</i>		
46. <i>Zea mays</i>		
47. <i>Pinus strobus</i>		
48. <i>Zea mays</i>		
49. <i>Pinus strobus</i>		
50. <i>Zea mays</i>		
51. <i>Pinus strobus</i>		
52. <i>Zea mays</i>		
53. <i>Pinus strobus</i>		
54. <i>Zea mays</i>		
55. <i>Pinus strobus</i>		
56. <i>Zea mays</i>		
57. <i>Pinus strobus</i>		
58. <i>Zea mays</i>		
59. <i>Pinus strobus</i>		
60. <i>Zea mays</i>		
61. <i>Pinus strobus</i>		
62. <i>Zea mays</i>		
63. <i>Pinus strobus</i>		
64. <i>Zea mays</i>		
65. <i>Pinus strobus</i>		
66. <i>Zea mays</i>		
67. <i>Pinus strobus</i>		
68. <i>Zea mays</i>		
69. <i>Pinus strobus</i>		
70. <i>Zea mays</i>		
71. <i>Pinus strobus</i>		
72. <i>Zea mays</i>		
73. <i>Pinus strobus</i>		
74. <i>Zea mays</i>		
75. <i>Pinus strobus</i>		
76. <i>Zea mays</i>		
77. <i>Pinus strobus</i>		
78. <i>Zea mays</i>		
79. <i>Pinus strobus</i>		
80. <i>Zea mays</i>		
81. <i>Pinus strobus</i>		
82. <i>Zea mays</i>		
83. <i>Pinus strobus</i>		
84. <i>Zea mays</i>		
85. <i>Pinus strobus</i>		
86. <i>Zea mays</i>		
87. <i>Pinus strobus</i>		
88. <i>Zea mays</i>		
89. <i>Pinus strobus</i>		
90. <i>Zea mays</i>		
91. <i>Pinus strobus</i>		
92. <i>Zea mays</i>		
93. <i>Pinus strobus</i>		
94. <i>Zea mays</i>		
95. <i>Pinus strobus</i>		
96. <i>Zea mays</i>		
97. <i>Pinus strobus</i>		
98. <i>Zea mays</i>		
99. <i>Pinus strobus</i>		
100. <i>Zea mays</i>		

clade: pteromalids

AUTHORS
Young, T. L.

TITLE **Development**

JOURNAL
PLANT PHYSIOL.
MEDITERR.
21563054

REFERENCE 2 (babe)

TITLE	HSP101 D
1. TITLE	
2. AUTHOR	
3. SUBJECT	
4. SUMMARY	
5. REFERENCES	
6. NOTES	
7. INDEXING	
8. EVALUATION	
9. CONCLUSIONS	
10. RECOMMENDATIONS	
11. REFERENCES	
12. NOTES	
13. INDEXING	
14. EVALUATION	
15. CONCLUSIONS	
16. RECOMMENDATIONS	
17. REFERENCES	
18. NOTES	
19. INDEXING	
20. EVALUATION	
21. CONCLUSIONS	
22. RECOMMENDATIONS	
23. REFERENCES	
24. NOTES	
25. INDEXING	
26. EVALUATION	
27. CONCLUSIONS	
28. RECOMMENDATIONS	
29. REFERENCES	
30. NOTES	
31. INDEXING	
32. EVALUATION	
33. CONCLUSIONS	
34. RECOMMENDATIONS	
35. REFERENCES	
36. NOTES	
37. INDEXING	
38. EVALUATION	
39. CONCLUSIONS	
40. RECOMMENDATIONS	
41. REFERENCES	
42. NOTES	
43. INDEXING	
44. EVALUATION	
45. CONCLUSIONS	
46. RECOMMENDATIONS	
47. REFERENCES	
48. NOTES	
49. INDEXING	
50. EVALUATION	
51. CONCLUSIONS	
52. RECOMMENDATIONS	
53. REFERENCES	
54. NOTES	
55. INDEXING	
56. EVALUATION	
57. CONCLUSIONS	
58. RECOMMENDATIONS	
59. REFERENCES	
60. NOTES	
61. INDEXING	
62. EVALUATION	
63. CONCLUSIONS	
64. RECOMMENDATIONS	
65. REFERENCES	
66. NOTES	
67. INDEXING	
68. EVALUATION	
69. CONCLUSIONS	
70. RECOMMENDATIONS	
71. REFERENCES	
72. NOTES	
73. INDEXING	
74. EVALUATION	
75. CONCLUSIONS	
76. RECOMMENDATIONS	
77. REFERENCES	
78. NOTES	
79. INDEXING	
80. EVALUATION	
81. CONCLUSIONS	
82. RECOMMENDATIONS	
83. REFERENCES	
84. NOTES	
85. INDEXING	
86. EVALUATION	
87. CONCLUSIONS	
88. RECOMMENDATIONS	
89. REFERENCES	
90. NOTES	
91. INDEXING	
92. EVALUATION	
93. CONCLUSIONS	
94. RECOMMENDATIONS	
95. REFERENCES	
96. NOTES	
97. INDEXING	
98. EVALUATION	
99. CONCLUSIONS	
100. RECOMMENDATIONS	
101. REFERENCES	
102. NOTES	
103. INDEXING	
104. EVALUATION	
105. CONCLUSIONS	
106. RECOMMENDATIONS	
107. REFERENCES	
108. NOTES	
109. INDEXING	
110. EVALUATION	
111. CONCLUSIONS	
112. RECOMMENDATIONS	
113. REFERENCES	
114. NOTES	
115. INDEXING	
116. EVALUATION	
117. CONCLUSIONS	
118. RECOMMENDATIONS	
119. REFERENCES	
120. NOTES	
121. INDEXING	
122. EVALUATION	
123. CONCLUSIONS	
124. RECOMMENDATIONS	
125. REFERENCES	
126. NOTES	
127. INDEXING	
128. EVALUATION	
129. CONCLUSIONS	
130. RECOMMENDATIONS	
131. REFERENCES	
132. NOTES	
133. INDEXING	
134. EVALUATION	
135. CONCLUSIONS	
136. RECOMMENDATIONS	
137. REFERENCES	
138. NOTES	
139. INDEXING	
140. EVALUATION	
141. CONCLUSIONS	
142. RECOMMENDATIONS	
143. REFERENCES	
144. NOTES	
145. INDEXING	
146. EVALUATION	
147. CONCLUSIONS	
148. RECOMMENDATIONS	
149. REFERENCES	
150. NOTES	
151. INDEXING	
152. EVALUATION	
153. CONCLUSIONS	
154. RECOMMENDATIONS	
155. REFERENCES	
156. NOTES	
157. INDEXING	
158. EVALUATION	
159. CONCLUSIONS	
160. RECOMMENDATIONS	
161. REFERENCES	
162. NOTES	
163. INDEXING	
164. EVALUATION	
165. CONCLUSIONS	
166. RECOMMENDATIONS	
167. REFERENCES	
168. NOTES	
169. INDEXING	
170. EVALUATION	
171. CONCLUSIONS	
172.	

REFERENCE 3 (base)

JOURNAL OF THE
DIRECTOR'S
SUBMITTEE

FEATURES

gene

CDS

670 a 874 c 1057 g 483 t
 SE COUNT
 ELVYDRDORSGGLVNAETGMMKSDJILLI QVPNSSTRSDAAQVCKMIEBDESDGNDES"
 TDAADLLISUSYDPVVGARPIRMVTEBRKPVQTSKMLIQBEIDENCTVYDIAAGPK
 RDLVNVREHRHFRPDLNRUDEIETDPUSHEQLRKVARLOKQVDAVRLAERGKSMKVA
 KAHVAVENTILLQVLDGRLTDGCGRTVDFRNVTVIMTSLGAHLLAGMVGKMSKVA
 FLDHNLVARDMESEVHQSARVLAGPGYVHGEGQLTEOVRVRETLALPDRVE
 RDRRNVQVTEVAAYAEVRLSRAGLRGPQPTGSPFLTGPTCVGTETLAKALARSL
 LQEDAAIKLESJETENKMLMTEVQPEQIAEVSRSQWTFIVTRLQMDNARLADVLKGLAD
 LDDRLKAIPTLMWKYKEXERIDEITRLKQORREELQTLQAEARMONARLADVLKGLAD
 AIDLVDACANVQJDSQFSEIDNLKRRKIQLEVELEHALKEKNDKASKARLITVRKE
 RFQVDFBACANVPDVSIVLGRKEQFEGHGVRQIDRWAAQLSARYIMGSHLPDK
 ILFIDEIHLVIGRAGTDSNDANFLKMLAGKQRCIGATTLLEYKVAIKVEBAGAKV
 VEGLAQRIQVPSNLLDVRVLIADMGALVAGQIRGEFPERLKAIVKVEBAGAKV
 TNPQALXTYGRDIAVEQAGKLDVPIGRDEEILRRVRLISRRTPKNNPVILGEPGVKTAI

SE CO
LIGN

Query Match 43.3%; Score 1345; DB 8; Length 3084;
 Best Local Similarity 69.1%; Pred. No. 0;
 Matches 1904; Conservative 0; Mismatches 835; Indels 18

141	GGATTACA	AAAGCTAATCG	AAGATG	CAAGAAATTC	ACACACA	GAAGCAACGAGAC	200	
142	GGAGTATG	CAAGAAAGACG	ACCATG	AAATCCG	GACCAATTC	CACCCA	GAAGCAACGAGGC	182
143	GGAGTATG	CAAGAAAGACG	ACCATG	AAATCCG	GACCAATTC	CACCCA	GAAGCAACGAGGC	182
201	AAATTGCT	CACAGCTCAT	GAGTAC	GTGAATCG	AGACATGCT	CAATTC	CACTCCCTTTGCA	260
183	GATCTG	TGGGGCGAC	GAGATTG	CGGTGG	AGCGCG	CACGCG	AGCTCAGCGCTGCA	242
261	TTT	PAGCTGGTCTT	TGATCTCT	GTATCC	CAACCGGTATATTTCT	CAAGCAATCTCT	TAGTGC	320
243	CCTGG	CGCAGTCTGG	CTGGCG	GACAAAGGGCG	GATCCTG	CGGACGCCATC	ACGGGGCG	302
321	CGTGG	CGGAGACG	CAGC-----	TC	AATCTGCTGA	AGAGTGATCAAT	CAAGCCTTGAA	374
303	GT	CGGGGGCG	ACGGAGCG	CGGGGACCTCG	TTGAGCGCGTGTCT	GAA	CAACTCGCTCAA	362
375	GAAGCTT	CTCTTCA	CAATCTCT	CCACCTCAT	GATATTC	CAGCGAGTTCT	TAGTCTTATTAA	434
363	GAAGCT	CTCGGTG	CAGTCTCC	CGCGCG	CGACCTCG	TC	CGGGCTCCAGGC	422
435	GGTCA	TTGCTGCTCA	AGCTGCTC	AGAAGTCA	CGAGGTGATATCT	CAATTTGGCTGTTGA	494	
423	GGT	CATCCGCGGGCG	AGTCCG	CGCAGAAAAA	CGCGGG	ACTCTCA	CCCTCGCGTGA	482
495	CGAGT	GATTATGGGTC	TTCTTGA	AGATTCTCA	AAATCAGGGATTTGTTGA	ACGAGAGTGG	554	
483	CCAGCT	GTGCTCGG	CTGCTCG	AGACTCG	CAGATCTCC	AGTCTCC	CAAGAGGCGCG	542
555	TGTAG	CGCAGCGG	TAAAGTCT	CAGGCTC	GAGAAGCTTC	GTGGGAAAGAGGGA	614	
543	CGTGT	TCGGCGCG	CGGTGG	CGCGAGCTT	GAGAAGCTC	CGCGCGGGGAGGGCGCG	602	
615	AGTTG	AGAGTGCTT	CAAGGGG	ACAAATTTT	CAAGCTTTAA	AGATTTA	TGAAGAGATT	674
603	CGTGG	AGTCCGCTCG	GGGGAC	CAACAACTT	CCAGGCGCTCA	AGACATAC	CGCGCGGACCT	662
675	GGTTG	AGCAGCAG	GGAGCTTG	ATCCTCTG	ATGCTGTATG	GTGATGAGGAGATTGA	AGAGT	734
663	CGTTG	AGCAGCGCG	GAAGCTG	GA	CCCGTCTATCG	CGCGCAGAGGAGATCC	CGCGCGT	722
735	CGTGAG	GATTTCTT	TCGAGGAG	ACGAAGAAC	CAATCCTGTGCTTATTTG	GAGAGCAGAGT	794	
723	GGTGC	GATCCTCT	CGCGCG	CAC	TAAAGAAC	CAACCCCGTCTCAT	CGCGAGCCCGCGT	782
795	TGTTAAA	CAGCTGTGGTT	GAGTTT	AGCA	CAAGAGATTTG	TGAAGAGGAGATGTG	CCCAA	854
783	TGG	CAAGACGCGCG	CTGTG	GAGGGCC	TCGCGAGCGCAT	CTGTT	TCGCGCGCAGCTGCCAG	842
855	CAGTCT	TACTGATGT	GAGATTAA	TTCTGTT	GACATGGGTGCGT	TAGTTGCTGGTGTCTAA	914	
843	TAACCT	CCTCG	AGCTCG	CCCTCAT	CGCTCG	ACATGGGCGCGCTG	TCGCGGGCGCAA	902
915	ATACCG	GAGGAGCTTT	GAAGAA	AGGTTG	AAATCTGTGTTTTGA	AAAGAAAGTTG	AGGACGCTGA	974

394 GCGCAGAAAGTCGCGCGGAGCTCGACACTCGCGCTCGACCAAGTGTCTCATGGGCTCTCTC 453
517 GAAGATTCTCAATCAGGGAATTTGTTGAAGAGTTCGGTGTAGCAGCGCGAGGTTAAAG 576
454 GAGGACCCGCGAGATCTCCAGCCGCTCAAGAGAGCGCGCATCTCCGCTGCGCGGTGAAG 513
577 TCTGAGGTTGAGAAGCTTCGTGGGAAAGAGGAAAGTTGAGAGTGTCTTCAGGGGAC 636
514 GCCGAGGTCGAGAAGCTCGGGGAGGCGAGCAACCGGCGGTGGATCTCGCGTGGGGAC 573
637 ACAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCGGGAAGCTT 696
574 ACCAACTTCAGGCCCTCAAGACGTACCGCCGCGACCTCGTGAGGTGCGCGCAAGCTG 633
697 GATCTGTGATTTGCTGCTGATCAGGAGATTAGAGTCTGTGAGATTTCTTCGAGGAGA 756
634 GACCGGTCTATCGCCCGCGCAGGAGATCCCGCGGTGGTCCGATCTCTGTCGGCGGC 693
757 ACGAAGAAACAATCTGTGTCTTATTTGAGAGCCAGGAGTTGGTAAACAGCTGTGTGTTGAA 816
694 ACAAGAAACAACCCCGTCTCATCGCGAGCCCGCGTGGGCAAGCCCGTGTGTGAG 753
817 GGTGTAGCAAAAGGATTTGAAAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTA 876
754 GGGCTCGCGCAGCGGTCTGCGCGCGCGACGTCCCGAGCAACCTCTCTGAGCGTGGCCTG 813
877 ATTTCTGTCACATGGTTCGCTTAGTGTCTGTCTTAATATCCGAGGAGAGTTTCAAGAA 936
814 GTCGCGTCTGACATGGCGCGCTCTGTGGCGCGCGCAAGTACCGCGCGAGTTCGAGGAG 873
937 AGGTTGAAATCTGTTTGAAGAAGTTGAGGACGCTGGAAGGCAAAAGTGAATCTCTTTATT 996
874 CGGCTCAAGCGCTGCTCAAGAGAGTGTGAGGAGCGCGAGGCGGTTCAGAGAGTGTATC 933
997 GATGAGATTCATTTGGTCTTGGTGTGCAAACTGAAGGCTGATGATGAGTACGCTAAT 1056
934 GAGGAGATACACTGGTGTCTGCGCGCGCGGAGCGAGGGGTCAATGAGCGCGCCAAAC 993
1057 CTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCTATGTTGCTACACGCTTGAA 1116
994 CTGTTCAAGCCCATGCTGGCGAGGGGCGAGTCTAGGTCATTTGGCGGACGACCTCGAG 1053
1117 GAATACAGGAATATTTGAGAAAGATGCTGCTTTTGAAGAGAGGTTCCAAACAGTCTAT 1176
1054 GAGTACAGGAAGTACGTCGAGAAGAGCGCGGTTTCGAGAGCGGTTCCAGCAGGTGTC 1113
1177 GTTGGGAGCAGTGTGCTGACACCATTAGTATCTCTAGGAGTCAAGGAGAGTAT 1236
1114 GTGGCGGAGCCAGCGTCCCGACACCGTCACTACTCAGAGGCTCAAGGAGAGTAC 1173
1237 GAGGACATCATGTTGCGAATCCAAGACAGAGCTCTTATAATGCTGCTCAGCTGCT 1296
1174 GAGGGGACCATGGGTGAGGATTCAGACCGCGGCTCTGCTATTTGCCGCGAGCTCTCG 1233
1297 GCTGTTACATAACTGGTGGCATTTACCGGATAAGCAATTTGATTTGGTTGATGAGGCT 1356
1234 TCGAGGTACATCATGGTTCGCTATTTGCTGATAAAGCAATCGATTTGTTGAGGAGGCC 1293
1357 TGTGCAATGTGAGAGTCCAGCTTGAATGATCACTGAGAGATTTGATTAACCTTGAAGG 1416
1294 TCGCAAAATGTGAGGTTGCACTTTCAGCAGCCGCTGAAGAGATTCGACCACTCGAGAGG 1353
1417 AAGAGATCAGCTGCAATTTGAACTTTCAGCGCTTTGAAAGGAGGAGATTAAGCCAGC 1476
1354 AAGAGATCCAGTTGAGAGTTCAGCTTCATGCTCTCGAAGAGGAGAGGACAGGCTAGC 1413
1477 AAGCTCGACTTATAGAGTGGGAAAGAGTGTGATGACCTGAGAGACAGCTTCAGCCT 1536
1414 AAGCTCGCTAGTTGATGTGAGGAAAGAGTTGGACGATCTGAGAGACAGCTCGAGCGC 1473
1537 CTCAGATGAATACAGAAAGAGAGAGAGATTTGATGATTCGAGGCTTAAACAG 1596
1474 CTGCAAGATGAGTATCGCAAGGAGAGAGAGGATTCGACGAGATCAGGAGCTTGAAGCAG 1533

QY 1597 AAAAGAGAGAGCTCATGTTTCTTTTCAGGAGCGAGAAAGAAAGATATGACCTTTGCAAGA 1656
Db 1534 CGCGCGAAGAGCTGCAAGTTTCACTTTCAGGAGCGCGAGCGCGATGATTTGGCCCGC 1593
QY 1657 GCTCTCATCTAAGATATGCGCAATTTCAAGAAAGTGAATCTGCAATTTGCCCACTTGAA 1716
Db 1594 GTGCTCATCTCAGATACGTTCTCTCAGAGAGTTGACGCTGCAATTTGTAAGCTGGAG 1653
QY 1717 GBAATCTTCTTGAAGAGAAATGTATGCTCAAGAAACGTTGGGCTTGAAACATTTGCT 1776
Db 1654 GG---TGAGACCGCGGAGAAATTTGATGTTAAAGAGACTGTGCGCGCGGACCAATTTGCT 1710
QY 1777 GAGGTTGTGAGCGGTTGAGCAGGATTTCCAGTACGAGACTTTGGCCAAATATGAGAGAG 1836
Db 1711 GAGGTTGTGAGCGGCTGACTGTTGTTCTGTCACAGGCTCGGACAGAAATGAGAGGCG 1770
QY 1837 AGGTTGATTTGCTCTTCTGATAGGTTGATAAGCGGTTGTGGGACAGAAATCAAGCGGTA 1896
Db 1771 AGGCTGATCGGCTGGCAGATCGACTGATCAGAGGTCGTTGGAAGATGAGGCGCTC 1830
QY 1897 AATGCAAGTTTCTGAGGCAATTTCTAAGTCAAGGCGAGGACTTTGGTAGGGGCAACAGCCA 1956
Db 1831 AATGCAAGTTTCTGAGGAGGCTGTTCTACGCTCGAGGCTGGCTTTGGAACGCGCGAGCGCT 1890
QY 1957 ACTGATCATTTCTTATTTCTTGGACCAACTGTTGGCAAACTGAGCTCGCAAGGCT 2016
Db 1891 ACTGTTTCTTCTTCTTGGACCGAGCGGTTGCGAAACCGAGCTCGCAAGGCT 1950
QY 2017 CTTGCTGAGCAGCTGTTTGTATGATGATAAACTCTTATGTTGCGATTTGATATGCGGAATAT 2076
Db 1951 CTAGCTGAGCAGCTGTTTGTATGAGAGAACTCTGCTCGCATCGACATGCTGTAATAC 2010
QY 2077 ATGGAACAACTCTGCTCTGCTCTGCTGCTCAATGGGAGCAACAGGAGTATGTTGTCAGGAG 2136
Db 2011 ATGAGCAGCATTTGCTGTTGCGCGCTAAATCGAGCGCCCACTCGGATATGTTGTCATGAA 2070
QY 2137 GAAAGTGGACAACTAACTGAGGCTGTGAGGAGCGAGCTTATTTGCTCATCTCTTTGAT 2196
Db 2071 GAAGGGGGGAGCTGAGCCGAGCAAGTGAAGAGAGGAGGCAATACAGTGTCTCTCTTCGAC 2130
QY 2197 GAAAGTGGAGAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAAAGTTTGGATGATGCT 2256
Db 2131 GAGTTTGAAGAGCGCATGTTGGCGGTTTCAACACTCTGCTCCAGGCTCTGAGCAGATGG 2190
QY 2257 CGATTGACAGACGCGGCAAGCGCAGACAGTCTGATTTCAAGAACTGCGGTGATATCATGACA 2316
Db 2191 CGGTTGACGACGCGGCAAGCGCAGGAGCGGTTGATTTTGAAGACACGCTGATCATGACC 2250
QY 2317 TCAAACTTGGTCTGTAACACTCTCTGAGGAGCTAACTGGGAAAGTAACTGGAAGTG 2376
Db 2251 TCAAACTTGGCGCGGAGCACTCTCTGCGGAGATGGTGGCA---TTTCGATGAGGTT 2307
QY 2377 GCCCGGAGCTGTGTGATGCGGAGGTGAGGAAACACTTTCAGACCAAGAGCTCTTGAACAG 2436
Db 2308 GCTCTGTGATCTGCTCATGAGGAGGTGAGGAGCAATTCGCGCGGAGCTGCTGAACCGT 2367
QY 2437 CTTTGAAGAGATTTGTTGTTGACCCCTTTTCACTACAGCTTGAAGGAGAGTACGCTGG 2496
Db 2368 CTGAGCAGATTCGTTCTTCTGACCCCTTCTGCGCATGAGCAGCTGCGAAGGTCGCTGG 2427
QY 2497 CTTTCAATGAAAGAGCTGCTGCTGCGGCTTGTGAAAGAGGAGTTGCTTTGCGAGTCACT 2556
Db 2428 CTTTCAATGAAAGATTTGCGAGTCTGCTCTTGGAGAGGCGGTTGCTCTGCGCGCTCAC 2487
QY 2557 GATGCTGCTTTGAGCTATATCTTTGGCAGAGATTTATGACCCCGTGTATGCTGCTAGGCT 2616
Db 2488 GACCGCGCTTGGAGCTCATCTGCTCACTGCTTTACGATCCGCTCTATGCGCGCAGGCA 2547
QY 2617 ATAAGGAGATGATGAGAGAGAGGTTGTCAGAACTCTCAAGATGCTTCAAGATGCTTGGTGG 2676
Db 2548 ATCCGAGATGGATCAGAGAGAGGATGATGACGAGCTCTCAAGATGTTGATCCGCGAG 2607

2677 GAAATCGATGAAATCTCACTGTTTACATAGATGCAG-----GGCTGCTGATCTTGTG 2730
 2678 GAGATCGACGAGAACTCCACGGTGTATCATGACGCTGGCGCCAGCAAGGACGAGCTGACC 2667
 2731 TACCGGTAGAAA---GTGAGGTCTAGTGGAGCTTCAACAGCGAAGAGTCAAGTGTG 2787
 2668 TATGGCGTCCACAGCAAGCAGGAGGCTGTGTGAACGCGCGACGCGGCCACCAAGTCCGACATC 2727
 2788 CTGATTCATATTCCTAACCGCGCCAAAGAGAGTGTGACGCTCAGCGCGGTGAAGAGATG 2847
 2728 CTGATTCAGGTCTTACGCGAGCTGTGTGGGGGCGATGCGCGCGCGGTGAAGAGATG 2787
 2848 AGGATCGAGGA 2858
 2788 AAGATCATGCA 2798

SULT 11
 332981
 CUS
 Oryza sativa heat shock protein (HSP101) mRNA, complete cds.
 AF332981
 CESSION
 AF332981
 AF332981.1 GI:18033449
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.
 1 (bases 1 to 3148)
 Agarwal,M., Sahi,C., Young,T., Agarwal,S., Mitra,S., Ganesan,K.,
 Agarwal-Katiyar,S., Gallie,D.R. and Grover,A.
 Isolation and Analysis of HSP101 from Rice
 Unpublished
 2 (bases 1 to 3148)
 Young,T.E. and Gallie,D.R.
 Direct Submission
 Submitted (02-JAN-2001) Biochemistry, University of California,
 Boyce Hall, Riverside, CA 92521-0129, USA
 Location/Qualifiers
 1. 3148
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 1. 3148
 /gene="HSP101"
 135..2873
 /gene="HSP101"
 /product="heat shock protein"
 /codon_start=1
 /protein_id="AAL57165.1"
 /db_xref="GI:18033450"
 translations="MNPDPNFTNEALVAHEHLESEAGHAQLPLHLVAALADKGG
 ILRQALSQSGDAGADPDSFRTVSVGALKLPSPSPSPSPSPASTALIKVIRPQSAQ
 KRGDSHLAVDQLGLLESLIDSLCKEAGVSAARVAERLEKRGEGKRVESASGD
 TNPQALTKYRDVLDLGLDLPIDVDEIRVRILSRTRKNNPVLIGEPGVGKTAV
 VEGALQIRVGRDVPVSLDVLRLDGLVAGAKYRGEFEERLXAVLKEVEEAEGKV
 ILFDIHLVLGAGRTGSGMDAANLKPMLRGQLRCIGATTLLEKRVKVEKDAAPER
 RFOQVFAEVSVDPTISILRGLKEKVEHGHVRIODRLVVAALASRYVIMGRHLDPK
 AIDLVEACANVRVQLDSQPEEDINLERKRIQLEVEHHALEKEKOKAKARLVEYKE
 LDRLQIKLPWKYRKERLEIRKIKQREELQFTQAEERRMDLARVADLKYG
 LQEDIVAIKSETENMLTETVPEQAEVVSRTGTPTVRLGNDKRELVLGLAD
 RLHQVRVQABAVSAVAELRAGLRGPPQPTGSLFLGPTGVKTEKLAKALAEQL
 FDDENLLVRIDMEYMEHVSARLIGAPGVGHEGGQLTEQVRRRPYSVILFEVE
 KAVHVAVENTLLQVLDLGRITDGGRTVDENTVIINTSLGAEHLILAGWGNKMKVA
 RDLVMEQVRRHFRPELLRLDELIVFDPLSHEQLKVARLQMDVAVRLAERGVLA
 TDAALVDLSLSVDPVGTGKRSKDLIQVPNGAATGSDAAQAVKMRIMDEDEQMDSE"
 ELAYRVDNRGLVNAETGQKSDLIQVPNGAATGSDAAQAVKMRIMDEDEQMDSE"
 644 a 931 c 1087 g 486 t

Best Local Similarity 68.3%; Pred. No. 0;
 Matches 1869; Conservative 0; Mismatches 848; Indels 21; Gaps 5;
 QY 163 ATGATTCAGAGAAATTCACACACAGCAACAGCAATTCAGTCAAGCTCATGAGCTA 222
 DB 135 ATGATTCAGAGAAATTCACACACAGCAACAGCAATTCAGTCAAGCTCATGAGCTA 194
 QY 223 GCTGTGAATGAGAGACATGCTCAATTCATCTCTTTGATTTAGTGTGCTTTGATCTCT 282
 DB 135 GCGTCGAGAGCGGGGCGCGAGCTGACGCGCTGACCTGCTGCTGCGCGCTGCGCGG 254
 QY 283 GATCCACCGGTATTTCTCAAGCAATCTCTAGTCCGGTGGCGAGAAACGACAG----- 337
 DB 255 GACAGGCGGGGATCTCCGGCAGCGATCTCGCAGCGCTCCGGCGGGGACGCGCGGCGG 314
 QY 338 -CTCAATCTGCTGAAGAGTCAATCAATCAAGCCCTTGAAGAGCTTCTTCAATCTCTCT 396
 DB 315 CCGGACTCGTTCGAGCGCGCTGAGCGGGGCGCTCAAGAAAGCTGCCCTCGACGTGCGG 374
 QY 397 CCACCTGATGATTTCCAGCGAGTTCCTAGTCTTATTAAAGGTCAATTCGTCGTGCTCAAGCT 456
 DB 375 CCGCGGACAGCGTGGCGGCTCCACGCGCTGATCAAGGTCAATCCGCGCGCGAGTCC 434
 QY 457 GCTCAGAGTCAAGAGTGAATCTCATTTGGCTGTTGACCAAGTTGATATGGGTTCTTCTT 516
 DB 435 GCGCAGAAAGAGCGGGGACTCCCACTCCGCTCGACCAAGCTGCTGCTGGGCTCTCTG 494
 QY 517 GAAATCTCAATCAAGAGTTCCTGCAAGCAAGTCTGCTGAGCGCGGAGGCTTAAAG 576
 DB 495 GAGGACTCGCTTATCTCCGACTGCTCAAGAGCGCGGCTGCTGCGGGCGGGTCCGC 554
 QY 577 TCTGAGGTGAGAGCTTCTGTTGAAAGAGGAGGAGGAGTTCAGAGTTCAGAGGAG 636
 DB 555 GCGGAGCTCGAGAGCTCCGTCGCGGAGGCGCGCAAGGTGAGTCCGCTCCGCGGAC 614
 QY 637 ACAATTTTCAAGCTTTAAGACTTTGAGACTTTGAGAGAGATTTGTTGAGCAAGCAGGAGCTT 696
 DB 615 ACCAACTTCCAGGCGCTCAAGACCTAGCGCGCGACCTCGTCGAGCAGCGCGGTAAGCTC 674
 QY 697 GATCTGTGATTTGCTGCTGATGAGAGATTAAGAGTTCGTGAGGATTTCTTTCAGAGAGA 756
 DB 675 GACCCGCTCATTTGCGCGGAGAGAGATCCGCGCTGCTGCTGCGATCTTGTCCGCGCG 734
 QY 757 ACAGAGAAACAATCTGTCCTTTATTTGAGAGCAGAGAGTTCGTTAAACAGCTGTGTTGAA 816
 DB 735 ACAGAGAAACAATCTGTCCTTATCTCGCGAGCGCGGCTGCGCAAGACGCGCTGCTGAG 794
 QY 817 GGTTTAGCACAAGAGTTCGAAAGAGAGTTCGCAACAGCTCTTACTGATGTGAGATTA 876
 DB 795 GGCCTCGCGAGCGATCGTGCAGCGGAGCGTCCAGCAACCTGCTCGAGCTCCGCTC 854
 QY 877 ATTTCTGTTGACATGGGTGCGTTAGTTGCTGCTGCTAAATACCGAGGAGAGTTCGAAGAA 936
 DB 855 ATCGCGCTCGACATGGGCGCGCTCGTCCGCGGCGCAAGTACCGCGCGAGTTCGAAGAA 914
 QY 937 AGTTTGAATCTGTTTGAAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGTTCGTTATT 996
 DB 915 CGGCTCAAGGCGGCTGCTCAAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGTTATT 974
 QY 997 GATCAGATTCATTTGCTTGTGCTGCTGCGCAAACTCAAGGCTCGATGAGTTCGAGCTAAT 1056
 DB 975 GACGAGATACACCTCGCTCCTCGCGCGCGGAGAGCGAGGCTTCATGAGAGCGCGCAAC 1034
 QY 1057 CTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCTTTCGTTGCTTCAACGCTTGA 1116
 DB 1035 CTGTTCAAGCCCATGCTCGCGAGGCGCGAGCTCCGCTGCTCAGCGCGCACGAGCTGAG 1094
 QY 1117 GAATCAGAGAAATATGTTAGAGAAAGATGCTGCTCTTTCAGAGAGGTTTCAACAGCTAT 1176
 DB 1095 GAGTACCGCAAGTACGTCGAGAGAGCGCGCGTTCAGAGAGGCTTCCAGAGAGTTC 1154
 QY 1177 GTTCGGAGCGCAAGTGTGCTGACCACTAGTATCTTCTTACAGGAGTTCAGAGGAGTAT 1236

Query Match 41.5%; Score 1289.2; DB 8; Length 3148;

1155 GTCCGGAGCCGAGCGTCCCGACACATCAGCATCTCCGAGGCTCAAGGAGAAGTAC 1214
 1237 GAGGACATCATGCTGTCGAAATCCAGACAGAGCTCTTATAAATGCTGCTCAGCTGCT 1296
 1215 GAGGGCCACACGGTGTGAGGATCAGAGCCCGCGCTGCTGCTGCGCGCAGCTCTCG 1274
 1297 GCTGCTTACATAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1356
 1275 GCGAGTACATCATGAGGAGCGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1334
 1357 TGTGCGAATGTGAGAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1416
 1335 TCGCGAAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1394
 1417 AAGAGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1476
 1395 AAGAGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1454
 1477 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1536
 1455 AAGGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1514
 1537 CTCAGATGAATACAG 1596
 1515 TTGAGATGAATACAG 1574
 1597 AAAAGAGAGAGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1656
 1575 GCGCGAGAGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1634
 1657 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1716
 1635 GTCCGCGACCTCAAGTACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1693
 1717 GGAATCTCTTGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1776
 1694 --GAGCAGAGCGGGGAGAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1751
 1777 GAGGTGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1836
 1752 GAGGTGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1811
 1837 AGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1896
 1812 AGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1871
 1897 AATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1956
 1872 AGCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1931
 1957 ACTGATCATCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2016
 1932 ACCGCT 1991
 2017 CTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2076
 1992 CTCGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2051
 2077 ATGGAACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2136
 2052 ATGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2111
 2137 GAAGTGAGCAACTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2196
 2112 GAAGTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2171
 2197 GAAGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2256
 2172 GAGTTCGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2231
 2257 CGATTGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2316
 2232 AGGTTCACCGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2291

QY 2317 TCRAACCTTGGTGTGAACACCTCTTCCAGGGCTAACTCGGAAAGTAAACAATGGAAGTG 2376
 Db 2292 TCRAACCTTGGTGTGAACACCTCTTCCAGGGCTAACTCGGAAAGTAAACAATGGAAGTG 2351
 QY 2377 GCCCGGACTGTGTGATGCGGAGGTGAGGAAACACTTCAGACCCAGAGCTCTTGAACAGG 2436
 Db 2352 GCCCGTACCTTGTGATGCGGAGGTGAGGAGGCACTTCGCTCTGAGCTCTTGAACAGG 2411
 QY 2437 CTTGACGAGATTTGTGTTTCGACCCCTTTTCATGACCCAGTTGAGGAAAGTAGTCTGG 2496
 Db 2412 CTCGACGAGATTTGTGATTTGATTCGATCCCTTCTCCAGCACTCAGGAAGGTTCGCGCT 2471
 QY 2497 CTTCAATGAAGACGCTGCTGCTGCGCTTCTGTAAGAGGAGTTCCTTTGGCAGTCACT 2556
 Db 2472 CTCGATGAAGACGCTGCTGCTGCGCTTCTGTAAGAGGAGTTCCTTTGGCAGTCACT 2531
 QY 2557 GATGCTCTTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGCTGTATGCTGTAGGCT 2616
 Db 2532 GACGCGGCTGCACTCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2591
 QY 2617 ATAAGAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2676
 Db 2592 ATCAGGAGGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2651
 QY 2677 GAAATCGATGAAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2730
 Db 2652 GAGATGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2711
 QY 2731 TACCGGCTAGAAA---GTGGAGGTCTAGTGGAGGCTTCAACAGGCAAGAGTCAAGTGTG 2787
 Db 2712 TACAGGCTGAGCAACCGCGGGGTTGGTGAACCGCGAGCGGGGAGGAGTCCGACATC 2771
 QY 2788 CTGATTCATATCTTAACCG---GCCAAGAGAGAGTGTGATGCTGAGCGGCTGAAGAG 2844
 Db 2772 CTCATCAGGTGCTCAACCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2831
 QY 2845 ATCAGGATCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2882
 Db 2832 ATGAGGATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2869

RESULT 12

AF097363
 LOCUS
 DEFINITION
 Triticum aestivum heat shock protein 101 (Hsp101b) mRNA, complete cds.

AF097363

AF097363.1 GI:4558483

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

1 (bases 1 to 3107)

Campbell, J.L., Klueva, N.Y., Zheng, H.G., Nieto-Sotelo, J., Ho, T.D.

and Nguyen, H.T.

Cloning of new members of heat shock protein HSP101 gene family in

wheat (Triticum aestivum (L.) Moench) inducible by heat,

dehydration, and ABA(1)

Biochim. Biophys. Acta 1517 (2), 270-277 (2001)

21240210

11342108

2 (bases 1 to 3107)

Campbell, J.L., Klueva, N.Y. and Nguyen, H.T.

Direct Submission

Submitted (05-OCT-1998) Plant and Soil Science, Texas Tech

University, Mail Stop 2122, Lubbock, TX 79409, USA

Location/Qualifiers

1..3107

/organism="Triticum aestivum"

/mol_type="mRNA"

gene
CDS
/cultivar="Mustang"
/db_xref="taxon:4565"
/dev_stage="10 day old seedling"
/note="treated 2 hrs. at 37 degrees C."
1. .3107
/gene="Hsp101b"
112. .2868
/gene="Hsp101b"
/note="Hsp101b"
/note="Hsp101b family member"
/codon_start=1
/product="heat shock protein 101"
/protein_id="A4D2629.1"
/db_xref="GI:4558484"
/translation="MNPDPNTHKNEALVAHEAASEAGHAQITPLHLAALAAADKSG
ILQAVAGAGNAGSAGSAPVPSASPASTALIKAIRAASQ
KRGDSHLAVDQLMLGLEDADIADCLKEAGVSARVRAELKLRGDSNRKVESASG
DINFOALYGRDLVRVAGKLDVIGRDEIRVRVRLIRKRNKPNVLPVIGEPGVOKTA
VEGLAQVRVGVPSNLLDVRVALDMGALVAGYRGEFEERLKAHLKAVLEVEBAEAG
VILFIDEIHLVAGRTSGMDAANLKPMLARGQLICIGATTLEBYRYKVEKDAEPE
RRQVVVAEAPSVAADTISILRGLKEKYEKGHVRIQDRAIVVAALISARYIMGRHLPD
KADLVDEACANVRVOLDOPERIDNLERKRIQLEVEHLALBKEKDKASKARLVVRK
ELDDLKDLQPLTMKYRKEKEIDEIRKLKORREELQFTLOEARRMDLARVADLKYG
ALGEIDALAKUGSTGENLMITETVPEQIAEVRVSRWTGIFVIRLQNDKRLVGM
DLHTRVGVTEAVNAVAEARSRLGRPOOPTGSLFPLGTGVKTELAKALAEQ
LFDENLLVRIDMEYMEQHSVARLIGAPPYVGHGEGQLTEQVRRRPYSVILFDEY
EKARVAVFTLLQVLDGRLTDQGRVDFRNTVIMTSLNLAELHLAGVKNKSVK
ARLVNMQEVRHPRPELLNRDLRIVVDFPLSHEOLKVARLQMKDVAVELRAGVALLA
VTDALDVILSLAYDPVYGARPRRWEKEVNTOLSKMLIQEIDENCTVYIDANKD
ELAYRVDRSGLVNATGORSRDLILQVNPALGGGKGGEAAKAVKMRVMEGDEEDGM
DEDV"

Query March 40.2%; Score 1246.8; DB 8; Length 3107;
Best Local Similarity 67.6%; Pred. No. 2.9e-204;
Matches 1851; Conservative 0; Mismatches 857; Indels 30; Gaps 6;

694 CTTGATCTCTGTGATTGGTGTGTGATGAGAGATTAGAAAGTCTGTGAGATTCTTTGAGG 753
652 CTCGACCCCGTCAATCGCGCGGACGAGAGATCGCGCGTGTGCGCATCTCTCAAGC 711
754 AGAAGCAAGAAACAATCTGTGTTTATTGGAGAGCCAGAGTGTGTAACAGCTGTGTT 813
712 CGCACCAAGAAACAACCCCGTCTCTCATCGCGAGCCCGCGTGTGGCAAGACGCGCGTGTG 771
814 GAAGGTTTAGCACAAAGATTGTGAAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGA 873
772 GAGGCTCTCGCGAGCGCGTGTGCGCGGAGACGTGCCCAACCTGTCTGAGAGTCCGC 831
874 TTAATTTCTGTTGACATGGGTGCGTGTAGTTTCTGCTCTAAATAACCGAGGAGATTGAA 933
832 CTGCTCGCGCTCGACATGGGCGCTGTGCGCGCGCAAGTACCGCGGAGTTCGAG 891
934 GAAAGTTTGAATCTGTTTGAAGAAGATTGAGACGCTGAGAGCAAGTGAATCTCTTT 993
892 GAGCGGCTCAAGGCGGTGCTCAAGGAGGTGAGAGGCGCGGAGGCAAGGTCATCTCTTC 951
994 ATTGATGAGATTCAATTTGGTGTCTGGCAAACTCAAGGGTGTGATGATGAGTCAAGT 1053
952 ATCGAGAGATACACTGCTCTCGCGCGCGGAGGAGCGAGGGTCCATGAGCGCGCC 1011
1054 AATCTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCAATTTGGTGTACAAAGCTT 1113
1012 AACCTGTTCAAGCGGATGCTCGCCAGGCGGCGAGTCCGGTGCATCGCGCGCACCCACGCTG 1071
1114 GAAGAAATACAGAAATATGTTGAGAAAGATGCTCTTTGAGAGAGGTTTCCAAAGTCT 1173
1072 GAGGAGTACCCCAAGTACGTCGAGAGGAGCGCGGTTTCGAGCGCGGTTTCCAGCAGGTG 1131
1174 TATGTTGCGGAGCAAGTGTGCTGACACCAATTAGTATCTTAGAGACTCAAGGAGAG 1233
1132 TAGTGGCGGAGCGAGCGTGTGCGCGACCACTACGATCTGCGAGGCTCAAGAGAG 1191
1234 TATGAGGACATCATGTTGCGAATCCAGACAGAGCTCTTATAATGTCTGCTGAGCTG 1293
1192 TACGAGGCGCACCGCGTGTGCGCATCGAGACCGCGCATCTGCTGCGCGCGAGCTC 1251
1294 TCTGCTGTTTACATACTGTTGCGCATTTTACCGGATTAAGCAATGATTGTTGATGAG 1353
1252 TCGGCGAGTACATCGGCGCGACCTGCGCGCAAGGCCATCGACCTGTTGGAGCAG 1311
1354 GCTTGTGCAATGTGAGAGTCCAGCTTATGATCAACTGAAGAGATTGATTAACCTTGAA 1413
1312 CGGTGCGCAACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1371
1414 AGGAAGAGATGAGTGTGAAATTAAGCTTCAAGCTTGAAGAGGAGAGATTAAGAGCC 1473
1372 AGGAAGCGGATTCAGCTGAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1431
1474 AGCAAGCTTCACTTATAGAGTGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1533
1432 AGCAAGCGCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1491
1534 CTTCTCAGATGAATAACAGAAAGAGAGAGATTTGATGAGTTCGAGGCTTTAAA 1593
1492 CGGCTGAGTGAATACAGGAAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1551
1594 CAGAAAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1653
1552 CAGCGCGCGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1611
1654 AGAGCTGTGATTAAGATATGCGCAATTCAGAAAGTGAATCTGCAATTTGCCCACTT 1713
1612 CGCTGCGCGAGCTTCAAGTACGCGCGCTCGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1671
1714 GAAAGAACTTCTTCTGAGAGAGATGTGATGCTCAAGAAAGCTTGGGCTTGAACACATT 1773
1672 GAGGG---CGAGCGCGGAGAACCTGATGCTCCACGAGACCGTCCGCGCGCGAGCATC 1728
1774 GCTGAGGTTGTGCGCGTGTGAGAGAGGATTCAGTGTGAGAGACTTGGCCCAATATGAGAG 1833

163 ATGAATCCAGAGAAATTCACACACAGACAAACGAGACAAATTTGCTACAGCTCATGAGCTA 222
112 ATGAACCCGACAACTTCACGCAACAGACCAACGAGCGCTGTGCGCGCGCACGAGCG 171
223 GCTGTGAATCAGACATGCTCAATTCATCTCTTTGATTTAGCTGCTGTTGATCTCT 282
172 GCGTGTGAGGCGCGCACGCGGAGATCACCGCGCTGCACCTGCGCGCGCGCTGCGCGG 231
283 GATCCCAACCGTATATTTCTCAAGCAATCTCTAGTGTGCGGTCGCGAGAA-----CGCA 336
232 GACAAGTGGGCACTCTGCGCGAGCGCTGCGCGGCGGCTGCGCGCGGAGACGCTCCGCG 291
337 GCTCAATCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAGCTTCTTCACAATCTCT 396
292 GGGGACTGTTGAGCGCGTGTCTCGCGCGCTCAAGAAAGCTGCGCTGCGAGTCCCC 351
397 CCAGCTGATGATATTCAGCGAGTTCTAGTCTTTTAAAGTCAATGCTGCTGCTCAAGCT 456
352 CCGCGGACTCCGTCGCGGCTTCACGCGCTCATCAAGGCCATTCGCGCGCGCGAGTCC 411
457 GCTCAGAAATCAGAGTGATCTCAATTTGGCTGTGACCAAGTGTGATTTAGGCTCTTCT 516
412 GCGCAGAAAGAGCGCGGAGTCCCACTGCGCTGCAACAGCTGTCTATGGGCTCTCTC 471
517 GAAGATTTCAAATCAGGATTTGTTGAACGAAGTGGTGTAGCGAGCGCGAGGGTAAAG 576
472 GAGGACGCGAGATCGCGCACTGCTCAAGGAGCGCGGCTGTGCGCTGCGGGTGGCG 531
577 TCTGAGGTTGAGAGCTTCTGTT---GGGAAAGAGGAGAAAGTTGAGAGTGTCTTCAAGG 633
532 GCGAGCTCGAGAGCTCCGCGCGCGGAGCAACTCGCGCAAGTTCGAGTCCGCTCCGCG 591
634 GACACAAATTTCAAGCTTTAAAGCTTATGAGAGAGATTTGGTTGAGCAAGCAGGAGAG 693
592 GACACCAACTTCCAGGCGCTCAAGAGCTACGCGCGAGCTCTGTCGAGGTTGGCGCGAG 651

1729 GCGAGGTGTGAGCCGCTGGACCGGCATCCCGGTACCCCGCTCGGGGAGAACACAG 1788
1834 GAGAGTTGATTGGTCTTCTGATAGTTGTCATAAGCGGGTTGTGGGACAGAAATCAAGCG 1893
1789 GAGCGCTCGTGGCATGCGGACAGGCTGCACACAGAGGTGTGGGACAGAGGCGC 1848
1894 GTAATGAGTTTCTGAGCAATCTTAAGTCAAGGCAAGGATTTGGTAGGCAACAG 1953
1849 GTGAACGCGCTGCGGAGCGGTGTGAGTCAAGGCGGGGCTTTGGGCGCGCAGCAG 1908
1954 CCAACTGATCATTTCTATTCTTGTGACCACTGGTGTGGCAAACTGAGTCTCCCAAG 2013
1909 CCAACGGGTGCTTCTTCTTCTGAGCCGACCGGCTGGGACAGCGAGTCTCCCAAG 1968
2014 GCTCTGCTGAGAGCTGTTGATGATGAACCTTTAGTTGCGATTGATATGCGGAA 2073
1969 GCGCTCGCGAGCAGCTGTTCGACGACGAGAACCTTCTGCTCGCATGACATGTCGAG 2028
2074 TATATGAAACACACTCTCTCTCTGCTCATGTTGGGACACACAGGATGTTGTCAC 2133
2029 TACATGAGCAGCACTCTCTGCTGCGCGCTCATCGAGGCGCACCTGGCTATGTGCGGCAC 2088
2134 GAGGAGGTGGACAACTAATCAGGCTGTGAGGCGGACCTTTATTTGTCTATCTCTTT 2193
2089 GAGGAGGCGGCGAGCTGACGAGCAGGTGAGGAGGCGGCTGACGCTGATCTCTGTC 2148
2194 GATGAGTGGAGAGCTCATGTTGCTGCTTCAACACTCTCTCAAGTTTGTGATGAT 2253
2149 GACGAGTGGAGAGCGCACGTGGCGGTGTTCAACACGCTCTCCAGGTGTGACGAC 2208
2254 GCTCATGACAGAGCGGCAAGCAGGACGAGTCTGATTTTCAAGAACTCGTGTATCATG 2313
2209 GCGCGCTGACGAGCGGCAAGCAGGAGGCTGAGTCTTCCGAAACACCGTGTATCATG 2268
2314 ACATCAACCTTGTGCTGAAACACTCTCTGAGGCTAACTGGGAAAGTAAACATGAA 2373
2269 ACCTCAACCTCGCGCGGAGCACCTCTCGCGGATGGTGGGCAAGCACTCATGAG 2328
2374 GTGGCGGAGCTGTGTGATGCGGAGGTGAGGAAACACTTCAGACCGAGCTTTGAAAC 2433
2329 GTCGCGCGGACCTCGTATGATGAGGAGGTGAGGAGGCACTTCAGCGCGAGCTGCTGAA 2388
2434 AGCTTGTGAGGATGTGTTGTTTCCAGCCCTTTCATGACGAGTGTGAGGAGTACCT 2493
2389 CGCTGTGAGGATGTGTTGTTTCCAGCCCTTTCATGACGAGTGTGAGGAGTACCT 2448
2494 CGCTTCAATGAAAGAGCTGCTGCTCGGCTTGTGAAAGAGGAGTGTGCTTTGGCAGTC 2553
2449 CGCTTGCAGATGAAGAGCTGCGCGCTGCGGCTCGCGAGCGGCGCTCGCCCTGGCGTC 2508
2554 ACTGATGCTGCTTTGACTATATCTTGGCAGAGTATGACCCGGTGTATGTTGCTAGG 2613
2509 ACCGAGCTGCGCTGGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2568
2614 CCTATAAGGAGATGATGAGGAGGAGTGTGACAGAACTGTCAAGATGTTGTTGCTGCT 2673
2569 CGATCAGAGGTGATGAGGAGGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 2628
2674 GAGGAAATCGATGAAATCTCACTGTTTACATAGATGAG ---GGCTGGTGTGATCTTGTG 2730
2629 GAGGAGATCGAGAGAACTGACCGGTGTACATCGACCGCGCCCAACAGGAGGCTGACC 2688
2731 TACCGGTGAGA ---AAGTGGAGTCTAGTGGAGCTTCAACAGGCAAGAGTGTGATGTTG 2787
2689 TACCGGTGAGACCGGAGCGGCGGCTGTGAGACCGGAGCGGCGGAGGCTGCGACATC 2748
2788 CTGATTCTATTTGCTAACGGGCGC-----AAAGAGAGTGTGACGCTCAGCGC 2835
2749 CTCATCCAGGTGCTCCCAACCGGCGCTCTCGGCGGCGGCAAGGCGGCGGCGGCGCAAGGCC 2808
2836 GTGAAGAGATGAGGATCGAGGAAATAGAAATGACGA 2873

Db 2809 GTGAAGAAGATGCGGGTCTATGAGGACGCGGACGAGGA 2846

RESULT 13
AF083344
LOCUS
DEFINITION
Triticum aestivum heat shock protein 101 kDa (HSP101) mRNA,
complete cds.
ACCESSION
AF083344
VERSION
AF083344.2 GI:11561807
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 2821)
Wells, D.R., Tanguay, R.L., Le H. and Gallie, D.R.
HSP101 functions as a specific translational regulatory protein
whose activity is regulated by nutrient status
Genes Dev. 12 (20), 3236-3251 (1998)
99003076
PUBMED
9784498
REFERENCE
2 (bases 1 to 2821)
Gallie, D.R., Tanguay, R.L. and Wells, D.R.
Direct Submission
Submitted (10-AUG-1998) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
REFERENCE
3 (bases 1 to 2821)
Gallie, D.R., Tanguay, R.L. and Wells, D.R.
Direct Submission
Submitted (06-DEC-2000) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
REMARK
Sequence update by submitter
On Dec 6, 2000 this sequence version replaced gi:3982832.
FEATURES
Location/Qualifiers
1..2821
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
1..2821
/gene="HSP101"
65..2821
/gene="HSP101"
/codon_start=1
/product="101 kDa heat shock protein"
/protein_id="AAC83689.2"
/db_xref="GI:11561808"
/translation="MNPDPNTHKNEALVAHEAASEAGHAIITPLHAAALAAKSG
ILKQAVAGSGGNAAGDSFEVLAGALKLPSQSPSPSVASTALIKAIRAQAQ
KRGDSHLAVDQLMGLLDAQIADCLKEAGVSASRVAELDKLRGDSNRKVESAFG
DTTFQALKTGRLDVEVAGKLDPIVDRDEIRRVIRILSRRTNNPVLIGEPGVGKTA
VHGLAQVRVGDVPSNLLDVLVADLMGALVAGAKYGEFEERKAVLKEVEEADGK
VILFIDEHLVAGAGTSGMDAANLKPMLARGQLRCIGATTLERYKYVEKDAFE
RRQGVVYKPSVADTISILGLKEKYGHGVRIQDRAIVVAOLSAKYIMGRHLPD
KADLVDEACANVRVQLDSQPEIDNLEKRIQLVELHALKEKDKASKARLVEVRK
ELDLKDLKPLTKRKERIDEIRKLQRRELQFTLQEARRRMDLARVLDLYG
ALQIEDAAIAKLEGETGENMLTETVGPQIAFVSVRWGTGIPVIRIGQNDKRLVGM
DLRHTVAVQTEANAVABAVLRSGALRGPQPTGFLGFTGVGKTELAKALAEQ
LPDENLLVIRIDMSYEVMSVARLIGAPGVVHEGGQLTEQVRRRPSYVILFDEV
EKHADVNTLLQVLDGSLTQDQGTVDPRNTVIMTSLNCAHLGAGWCKNSMKY
ARDLVNQEVRRHFRPELLNRLDENWIFDLSHEQLRKVARLQMKOVAVRLAERGVALA
VTDAALDVLISLAYDPVYGARFIRWIEKRVVTLQSKMLIQEIDENCTVYIDAAKD
EDV"
BASE COUNT 524 a 902 c 1035 g 360 t
ORIGIN

Query Match 39.8%; Score 1236.2; DB 8; Length 2821;
Best local Similarity 67.3%; Pred. No. 1.4e-301;
Matches 1847; Conservative 0; Mismatches 868; Indels 30; Gaps 6;

QY 156 ATCGAAGATGAATCCAGAGAAATTCACACAGACAAACGACAGCAATTTGCTACAGCTCA 215

1138 GGAGAAGTACAGAGGGGACACACCGCGTGGCGATCCAGACCGCGCCATCTGCTGTCGCGGC 1139

1287 TCAGCTGTCTGCTCGTTTACATAAATCTGTCGGCAATTTACCGGATAAAGCAATTGATTTGGT 1346

1198 GCAGCTCTCGGCGAGGTACATCATGCGCGCGCCACTCTGCGGACAAGGCGCATCGACTGGT 1257

1347 TGATGAGGCTTGTCCGAATCTGAGAGTTCAGGCTTGATAGTCAACTCTGAAGAGATTTGATAA 1406

1258 GGCAGGCGGTGCGCCAACTGAGGGTGCAGCTTGCACGCGCGGAGGAGATTCGACAA 1317

1407 CTTTGAAAGGAAGGAGTGCAGCTGGAAATTTGAACTTTCACGCTTTGAAAGGAGGAAGGA 1466

1318 TCTGAGCGGAAGCGGATTCAGCTGGAGGTGGAGCTTGACGCGCTGGAAGAGGAGAAGGA 1377

1467 TAAAGCCAGCAAAAGCTCGACTTATAGAGTTCGGAAAAGAGCTTGATGACCTTGAGAGCAA 1526

1378 CAAAGCCAGCAAGCGCCGCTGGTGGAGGTGAAGAGGAGTGGACGACCTTGAGGGCA 1437

1527 GCTTCAGGCTCTCACGATGAAATACAGAAAGAGAAAGAGAGATTTGATGAGATTCGAAG 1586

1438 GCTGCAGCGCTGACCATGAATACAGGAGGAGGAGGCGGATTCGACGAGATCCGGA 1497

1587 GCTTAAACAGAAAAGAGAGCTCATGTTTTCTTTGAGAGGCGCAGAACGAGATTTGA 1646

1498 GCTGAAGCAGCGCGGAGGAGCTGCAGTTTTCAGCTGCAGGAGGCGCGAGCGGCATGGA 1557

1647 CTTTCAAGAGCTGCTCATCTAAGATATGCGCAATTTCAAGAAGTGGAAATCTGCAATTTGC 1706

1558 CTTGCGACGCTGCGCGACTCAAGTACGCGCGCTTCAGAGAGTTCGACGCGCCATTTGC 1617

1707 COAATTGAGGAATCTTCTTTCGAGAGAAATGTGATGCTTACAGAAAAGTTGGCGCTGA 1766

1618 GAACTGTGAGGG---CGAGACCGCGAGAACTGATGCTTACCGAGACGCTGCGCGCCGA 1674

1767 ACACATTCCTCAGGTTGTGAGCGTGTGGACGAGATCCAGTGACGAGACTTTGSCAAA 1826

1675 GCGATCGCGAGGTGTGAGCGGTGGACCGGCAATCCCGGTGACCCGGCTTCGGGCGAGA 1734

1827 TGAGAGGAGAGGTTGATTTGGTCTTGCTGATAGTTTGCATAAGCGGGTTTGGGACAGAA 1886

1735 CGACAAGAGCGCTGCTGGCATGGCGGACAGGCTGCACACAGGGTGGTGGCGGACAG 1794

1887 TCAGCGGTAAATCGAGTTTCTGAGGCAATTTAGGTCAGGCGCAGACTTGTGTAGGCG 1946

1795 GGAGCCCGTGAACCGCTCGCCGAGCGGTGTGAGGTTCCAGGGCGGGGCTTTGGCGCGCC 1854

1947 ACACAGCCAACTCGATCATTTCTTTCTTGACCAACTGGTGTGTGCAAAACTGAGCT 2006

1855 GCAGCAGCCACCGGTCGTTCTCTTCTGGGCCGACCGGCGTGGGCAAGACAGAGCT 1914

2007 CGCCAAGGCTCTTCTGAGCAGCTGTTTGATGATGAAACCTCTTAGTTTCGATTTGATAT 2066

1915 CGCCAAGGCCCTCGCCGAGCAGCTGTTTGACACGAGAACCTTGCTCGTCCGATTCGACAT 1974

2067 GTTCGAATATATGAAACAACTCTGCTCTCGCTCATTTGGGCGACCCACCGAGGTATGT 2126

1975 GTCCGATGATGAGCAGCATCTCGTTCGCTGCTTCACTTCGCGCGCCACCGAGCTATGT 2034

2127 TGGTCA CGAGGAAGGTGGACAACTAACTGAGGCTGTGAGGCGGACCTTTATTTGTGTAT 2186

2035 CGGACACGAGGAGCGGCGAGCTGACGAGCAGAGTGTGAGGAGGAGCGGCTACAGCGTGAT 2094

2187 ACTCTTTGATCAAGTGGAGAGGCTCATGTTGCTGTCTTCAACTCTGCTCCAGTTTTT 2246

2095 CTTGTTTCAGCAGGTGGAAGGCGCAGCTGGCGGTGTTTCAACAGCTCTCTCCAGGTGCT 2154

2247 GGAATGATGGTTCGATTTGACAGCGGCAAGGCGAGGACGTTCGATTTTCAGGAACCTCGGTAT 2306

2155 GGAACGCGCGGCTGACCGACGCGCAAGGCGGACGGTGTGACTTCGGAACACCGTGAT 2214

2307 AATCATGACATCAAACTTTGGTGTCTGAAACACTCTCTCTGACGGGTAACTGGGAAAGTTAAC 2366

2215 CATCATGACTTCAACTCTGGCGCCGAGACCTCTCTCGCCGATGTGGGCAAGAACTC 2274

2367 AATGAAGTGGCCCGGAGCTGTGTGATCGCGAGGTGAGGAAACACTTCAGACCAAGACT 2426
|||||
2275 CATGAAGTGCCTCCGCGAGCTCTGTATGACGAGGTGAGGAGGCACTTCAGCGCGAGCT 2334
|||||
2427 CTTGAACAGGCTTGACGAGATTGTGTGTTCGACCCCTTTCCACATGACCACTGAGGAA 2486
|||||
2335 GCTGAACCGCTGGACGAGATGTGTATCTTCACCCGCTGTCCACGAGCAGCTGAGGAA 2394
|||||
2487 AGTAGTCGGCTTCAAAAGAGAGCTGTCTCGCGCTTGTCTGAAAGAGAGAGTTGCTTT 2546
|||||
2395 GGTGGCCCGCTCGAGATGAAGAGAGTGTGCGCTCGCGCTCGCGAGCGGCGTGCCT 2454
|||||
2547 GCAGTCACTGATGCTGCTTTGGACTATATCTTGCAGAGATTATGACCGGTGTATGG 2606
|||||
2455 GGCCTGTACCGGCTGGACCGCGCTGAGCTCATCTGTCTGCTGGCTACGATCCAGTGTATGG 2514
|||||
2607 TGCTAGCGCTTATAAGGAGATGATGAGAGAGAGTGTGTGACAGAACTGTCAAAGATGCT 2666
|||||
2515 CGCTCGCGCGATCAGAGGTGATCGAGAGAGGTGTGTGACGAGCTGTCCAGATGCT 2574
|||||
2667 TGTGCTGAGGAAATCGATGAACCTCCACTCTTTACATGATGACGCG---CTGGTGA 2723
|||||
2575 GATCCAGGAGGAGATCGACGAAATTGCACTGTGTACATCGACGCGCGGACAGGACGA 2634
|||||
2724 TCTTGTGTACCGGTAGA---AAGTGGAGGTCTAGTGGACGCTTCAAAGGCAAGAGTC 2780
|||||
2635 GCTGGCTTACCGGCTGGACCGCGCGGCTGTGTGAACCGCGGAGACGCGGCGAGGTC 2694
|||||
2781 AGATGTGCTGATTCTATTTGCTTAACGGGCGC-----AAGAGAAAGTGTGACG 2828
|||||
2695 GCACATCTCATCGAGTGCCCAAGCGGCGCTCGCGCGCGCGCGCGGCGAGCGGC 2754
|||||
2829 TCAGCGGTGAAGAGATGAGATCGAGGAAATAGAGATGACGA 2873
|||||
2755 CAAGCGCGTGAAGAGATGAGGCTCATGAGGACGCGGACGAGGA 2799
|||||

ULT 14

HRIV39

US

INITIATION

SESSION

AL161539

AL161539.2

GI:7268147

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 197976)

EU Arabidopsis sequencing project.

Direct Submission

Submitted (10-MAR-2000)

MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:

lenckemips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project

Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge

Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

E-mail: michael.bevan@bbsrc.ac.uk

Information on performance of analysis and a more detailed

annotation of this entry and other sequences of chromosomes 3, 4

and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>

this fragment has an overlap with ATCHRIV40 at the 3' end and an

overlap with ATCHRIV38 at the 5' end.

Location/Qualifiers

1..197976

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

5668..7235

/gene="AT4g14420"

gene

CDS

complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))
/gene="AT4g14420"
complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))
/gene="AT4g14420"
/notes="strong similarity to Nicotiana tabacum ORF able to
induce HR-like lesions", PATX:G1762945
contains EST gb:AI999161.1"
/codon_start=1
/product="elicitor like protein"
/protein_id="CAB78484.1"
/db_xref="GI:7268148"
/translation="MELASFLGRALFVSVLLSAWQFNDFGEDGGRSAKSLKPFNA
FVNVHTHTQQQLPPVDMKILVAALAKIGLLFVFGSSLGAYLLILHQAATPIL
YDFVYDVRKRFQSLKFTQSLALLGLLPIFGMKNRKHGRQLRKKAPKAKN"
complement(5668..5772)
/gene="AT4g14420"
/number=1
complement(5773..5864)
/gene="AT4g14420"
/number=1
complement(5865..5963)
/gene="AT4g14420"
/number=2
complement(5964..6050)
/gene="AT4g14420"
/number=2
complement(6051..6140)
/gene="AT4g14420"
/number=3
complement(6141..6237)
/gene="AT4g14420"
/number=3
complement(6238..6352)
/gene="AT4g14420"
/number=4
complement(6353..7167)
/gene="AT4g14420"
/number=4
complement(7168..7235)
/gene="AT4g14420"
/number=5
8407..9129
/gene="AT4g14430"
complement(8407..9129)
/gene="AT4g14430"
complement(8407..9129)
/gene="AT4g14430"
/notes="similarity to carnitine racemase, Escherichia coli
contains EST gb:H36845, AI992737.1, T76583"
/codon_start=1
/product="carnitine racemase like protein"
/protein_id="CAB78485.1"
/db_xref="GI:7268149"
/translation="MCTLEKRGDLFLTLTGDEHRHPDPTIATLSLLEQXQSTR
GSLITANGKFFSNGFDLAWTAGSKTGAANRLHQMVSFQVVAALLDLPMTIA
ALNGHAAAGLILALSHDYVFNKRDGVLYMSEVDIGLSMPDYFSAIVRAKIGTSAAR
RELLSGKKIRGEAVGLGIVDSAAVSEEGVVAVSLRGLERKLAQKWSGEVYASIRK
SLYPELCGLGLETRVPATPKL"
complement(8407..9129)
/gene="AT4g14430"
/number=1
10443..11159
/gene="AT4g14440"
complement(10443..11159)
/gene="AT4g14440"
complement(10443..11159)
/gene="AT4g14440"
/notes="similarity to carnitine racemase, Escherichia coli
contains EST gb:AI994308.1, AA394352, H36791, AA395864,
T76583, H77091, T46265, AA597888, Z17630"
/codon_start=1

FEATURES

source

	DQIKAGNLIKALHVPTEHQADILFKALHPGPFHHLLKQMSLSLFLPNKPVR*
exon	17575..19093 /gene="AT4gl4460" /number=1
intron	19094..19144 /gene="AT4gl4460" /number=1
exon	19145..19476 /gene="AR4gl4460" /number=2
intron	19477..19551 /gene="AT4gl4460" /number=2
exon	19552..21256 /gene="AR4gl4460" /number=3
intron	21257..21337 /gene="AT4gl4460" /number=3
exon	21338..21472 /gene="AT4gl4460"
Query Match 31.5%; Score 978.8; DB 8; Length 197976; Best Local Similarity 65.8%; Pred. No. 2.3e-236; Matches 1724; Conservative 0; Mismatches 577; Indels 319; Gaps 8;	
Qy	456 TGCCTCAGAAGTCACGAGGTGATACTCATTTGGCTGTGTGCCAGTTCGAATTATGGTGCTTCT 515
Db	114161 TGCAAAAAGAAATCTCGGGGTACC AAAAGTGGGAGTAGCGGTATTGGTAATTAGTCTTCT 114220
Qy	516 TGAAGATCTCAAATCAGGATTTTTGTGAACGAAGTCGGTGTAGCCAGCGCGAGGGTAAA 575
Db	114221 TGAAGATCTCAAATCAGTGATGTGTGAAGAAGACTGGTGTGTGCTTGAGAAGGTAA 114280
Qy	576 GTCTGAGGTGTGAAGACTTCTGTGGAAGAAGGAAGAAAGTTGAGAGTGCTTCACGGGA 635
Db	114281 GTCTGAGGTGGAGAAGCTTCTGT-----GGGGA 114307
Qy	636 CACAATAATTTCAAGCTTTAAAAAGATTATGGAAGAGATTTCGTTTGACCAAGCAGGGAAGCT 695
Db	114308 GGTAATCTTCGGGCTTTGAAGACTTATGGCACAGATTTCGTTTGACCAAGCAGGTAACT 114367
Qy	696 TGATCTGTGATGTGTCGTGATGAGAGATPAGAAGAGTCGTGAGGATTCCTTCGAGAGAG 755
Db	114368 TGATCTGTGCATTTGCCACACACAGCGAGATTAGAAGAGTAATTTGAGGTTCTCTCAAGGAG 114427
Qy	756 AAGCAAGAACAACTCTGTGCTTTATTGGAGAGCCAAGGATGTGPTAAACAGCTGTGTTGA 815
Db	114428 AACCAAGAACAAACCCGTCCTCATCGGTGAGCCAGGTGTTGGAAAACCTGCTGTGTTGA 114487
Qy	816 AGGTTTACCAAAAGGATTTGAAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATT 875
Db	114488 AGGTTTGGCTCAAAGGAFACTGAAAGGAGACGTGCCCTATCAATCTTACTGTTAAAACT 114547
Qy	876 AAATTCGTTTGACATGGGTGCGTTAGTTGCTGCTGTAAATACCGAGGAGAGTTTCAGAG 935
Db	114548 TAATTCCTTGGAAATTTGGTGCAATPGGTGTCTGGAAACAACACTCCGTTGGCAATTCGAGA 114607
Qy	936 AAGGTTGAAAATCTGTTTTGAAAGAAGTTTGAGGACGCTGGAAGGCAAAAGTGATCTCTTTAT 995
Db	114608 AAGATTGAAGTCTGCTTGAAAGCAGTTTGAGGAAGCTCAAGGCAAAAGTTGTTCTGTTAT 114667
Qy	996 TGATGAGATTCAATTTGGTCTTGTGTCGTGCCAAACCTGAAGGGTCGATGGATGCAAGCTAA 1055
Db	114668 CGATGAGATCCCAATGGCTCTTGTGTCGTTTGAAGAAGGAGTGGGTCAACGGATGCTGCTAA 114727
Qy	1056 TCCTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCAITTTGGTGTCTACAAACCTTGA 1115
Db	114728 GTTGTCTAAGCTATGTTAGCAAGGTCAAGTTCAGTTCGTTTCATTGAGACTACTACGCTTGA 114787
Qy	1116 AGAATACAGGAAATATGTTTGAAGAAGATGCTGCTTTTGAGAGAGGTTCCAA CAAGTCTTA 1175
Db	114788 GGAAATACCGAACACATGTTTGAAGAAGATGCTGCTTTTGAGAGAAGGTTCCACCAAGTGT 114847

1176 TGTTGGGAGCCAAAGTGTGCTGTGACACACATTAGTATCTTAGAGGACTCAAGGAGAGTA 1235
 114848 TGTAGCTGAGCCTAGTGTGCTGATGATCTATCAGTATCTTCGAGGGCTTAAAGAGAAATA 114907
 1236 TGAGGACATCATGCTGTGCGAATCCAGACAGAGCTTATAAATCTGCTCAGCTGTC 1295
 114908 TGAAGGACATCATGCTGTGCGAATCCAGACAGAGCTTATAAATCTGCTCAGCTGTC 114967
 1296 TGCTCGTTACATAACT 1311
 114968 TGAACGTTTACATCACTGCTATGTTGCGCCATCTTTTATCTAACATGCTCTTGTAAATTA 115027
 1312 115028 TTATGATCTTTTATGTTGCTTATTAGTGTCTGTGCTGTAGTCTGAGCTTACC 115087
 1326 GGATAAAGCAATTCATTTGTTGATGAGGCTTGTGCGAATCTGAGAGCTCAGCTGTAG 1385
 115088 GGATAAGCAATTCATTTGTTGATGAGTCTTGTGCAATCTCAAGCCAGCTTGATAT 115147
 1386 TCAACCTGAAGAGATTGATAACCTTGAAGAGAGAGGATGAGCTGGAATTTGAACCTCA 1445
 115148 TCAACCAAGAGAAATTCATAGCTTTGAAGAGAGAGTAAATGCACTAGAGATAGATTCA 115207
 1446 CGCTTGGAGAGGAGAG---GATAAGAGCAGCAAGCTCGAC 1486
 115208 TGCTCTCGAAGAGAGAGAGAGAGAGCTAGCTGCTTCTCCGAGGTAAATTA 115267
 1487 115268 CTTTCAAAACATGCTTTTACATATAGTACCTCACTCTTTTATTTGTTATTTCTGCAGG 115327
 1496 TGGGAAAGAGCTTGTGACCTGAGAGACAGCTTCAAGCTCTCAGCATGAAATACAGAA 1555
 115328 TGAGAAAGAGCTTGTGATCTTGAAGGAGAGCTTGAAGCTCTTACAAATTAATACAAA 115387
 1556 AGGAGAAAGAGAGAAATTCATGAGATTCGAGGCTTAAACAGAAAGAGAGAGCTCATGT 1615
 115388 AGGAGAAAGAGAGAAATTCATGAGATTCGAGGCTTAAACAGAAAGAGAGAGCTCATGT 115447
 1616 TTCTTTTGGAGGAGAGAGAGATATGACCTTGCAGAGCTGCTGATTAAGATATG 1675
 115448 TTGCTCTACAGAGGCTGAGAGACATGACGTTTCTAAAGCTGCTTCTAAATATG 115507
 1676 GCGCAATTCAGAGAGTGAATCTGCAATTCGCCAACTTGAAGCAACTTCTTGAAGAGA 1735
 115508 GTGCAATTCAGAGAGTGAATCTGCGATAGCGAACT---AGAAAGAGTCTTAAAGATA 115564
 1736 ATGTGATCTCACAGAAACGTTGGGCTGAAACATCTGCTGAGGTTGTGAGCCGTTGGA 1795
 115565 ACGTGTGCTTACAGAAACTGTTGGGCGAGAGATATTTGCTGAGTGTGAGCCGATGGA 115624
 1796 CAGGATTCAGAGTGCAGAGACTTGGCCAAATCAGAGAGAGAGGTTGATTTGCTTGTG 1855
 115625 CTGGATACAGTGTGACTGAGTGTGACAGAAACGAGAGAGAGAGCTTAAATCTTGTG 115684
 1856 ATAGTTGCAATAAGCGGTTGTGGGACAGAAATCAAGCGTAAATGCAAGTTCTTGAGGCAA 1915
 115685 ATAAGTTGACGAGAGAGTGTGGGACAGAGCTGTGAAAGCTGTGGCTGTGCTGCA 115744
 1916 TTCTAAGTCAAGGCGAGACTTGTGTAGGCAACAGCCCACTGAGATCATTTTATTC 1975
 115745 TACTAAGGTCAAGGTTTGGACTAGGAGGCGCAACAAACCGCTGCTTCTCTCTCC 115804
 1976 TTGACCAACTGTTGTTGGCAAACTGAGCTCGCAAGCTCTTGTGAGCAGCTGTTG 2035
 115805 TCGGTCCAACTGTTGTTGGTAAACCGAGCTGAGCTGAGCTGAGCAACTTTG 115864
 2036 ATGATGAAACCTCTTAGTTTCGATTGATATGTCGGAATATATGGAACCAACACTCTGTCT 2095
 115865 ACAGTGAAACCTTACTCGTTAGGCTTGATATGTCGAGTATATGACAAATTTCTCTGTTA 115924

2096 CTCGCTCATTTGGGCGACCAACAGGCTATG 2125
 115925 ATAAACTTATAGGGCGACCACTGGATAGCTCTTAAACTCAAGCCACTTCTTTATGTGT 115984
 2126 2126
 115985 TTTTGTGATTAATGGTATATATGGAACCTCACATGATCTACTTTTGTCTTTTCTTGTAGTACAT 116044
 2127 TGTCTACAGAGGAGGTGGCAACTAACTAGAGCTGTGAGGAGCGACCTTATTTGTGTCTAT 2186
 116045 TGTCTATGAGAGAGGTGGCAACTAACTAGAGCTGTGAGGAGCGACCTTACTGTGTGT 116104
 2187 ACTCTTTGTATGAAGTGGAGAGGCTCATGTTGCTGTCTCTTCAACACTCTGCTCCAAAGTTT 2246
 116105 ACTATTTGATGAAGTTGAAAGACCCATGTTACTCTTATCAACAGCTACTTCAAGTCT 116164
 2247 GATATGATGCTCATTTGACAGAGCGGCAAGGAGGAGGAGCTGATTCAGAACTCGGTGAT 2306
 116165 AGAAGACGGGAGATTAAACCGATTAGCATGGGAGGAGCAGTCGATTTTAAAGAACACAGTGT 116224
 2307 AATCATGACATCAAAACCTTGGTGTGTAACACCTCTCTGAGGGCTAACTGGGAAAGTAAAC 2366
 116225 CATATGATCTTCTAACTTGGTGTGAGACCACTTGTTCAGGGCTTACAGGTGAATGAC 116284
 2367 AATGAAAGTGGCCCGGACTGTGTGATGCGGAGGTGAGGAAACACTTTCAGACCAAGCT 2426
 116285 AATGCAAGTGGCTCGAGACCAACGCAATGAAGATGCGAAGAAACATTTTCAGCCAGAGCT 116344
 2427 CTTGAAACAGGCTTGCAGAGATTTGTTGTTGACCCCTTTTCAATGACCAAGCTTGAAGAA 2486
 116345 GTTAAACAGATTTGAGATGATGTTATGTTTTCATCTTTTGTCCACGAGCATTTGGCAAA 116404
 2487 AGTAGCTCGGCTTCAAAATGAAAGAGCTTGTGCTCGGCTTGTCTGAAAGAGGAGTTGCTTT 2546
 116405 AATTGTTCACTTCAAGTGAGATGTTGCT 116437
 2547 GGCAGTCTACTGATGCTGCTTTGGACTATATCTTGGCAGAGAGTTATGACCCGCTGTATGG 2606
 116438 ---TGCTTGAAGTTTACATTTTATAATATGATGTTTATTTAGTGTATGG 116492
 2607 TGCT---AGGCTTATAGGAGATGATGATGAGAGAGGTTGTTGACAGAACTGTCAAAGT 2664
 116493 TGCTAGAGGCTTATTAAGATGCTGAGAGAGAGAGTGGTCAAGATATATCAATGATG 116552
 2665 GTTGTGCTGAGGAAATCGATCAAACTCCACTGTTTACATAGATGACGCGCTGTGAT 2724
 116553 ATTGTGCGAGAGAAATCGGTGACGATTCATTTGTCATGATGTTAAAGAGGCTAAA 116612
 2725 CTTGTGTACCGGCTAGAAAGTGGAGGCTTAGTGACGCTTCAACAGGCAAGAGTCAAGT 2784
 116613 ACTGATCTTGTCTACCGAATAGATAAAATGTTTTTGTGAAGATAGAGCAACATTAGAT 116672
 2785 GTGCTGATTCATTTGCTTAAACGGCCCAAGAGAGTGTATG 2824
 116673 GTTGTGATTCATTTGCGGATTAAGAGAGAGAGAGCAATG 116712

RESULT 15
 Locus ATFCAL
 DEFINITION Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
 No. 1.
 accession 297336
 version 297336.1 GI:2244788
 keywords
 source Arabidopsis thaliana (thale cress)
 organism Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N.,
 Kreis, M., Kavanagh, T., Entian, K.D., Rieger, M., James, R.,

Puigdomenech, P., Hatzopoulos, P., Obermaier, B., Duesterhoft, A.,
 Jones, J., Palme, K., Ansgore, W., Delany, M., Bancroft, I.,
 Mewes, H.W., Schueller, C. and Chaiwatzis, N.
 Unpublished
 2 (bases 1 to 206606)
 EU Arabidopsis sequencing project.
 Direct Submission
 Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 schueller@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with AtFCA0 at the 5' end and an
 overlap with AtFCA2 at the 3' end.

FEATURES

source	Location/Qualifiers	
	1..206606	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/variety="Columbia"	
	/db_xref="taxon:3702"	
	/chromosome="4"	
source	1..26121	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="BAC TAMU8H22"	
	12367..32771	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1N11"	
	32771..36017	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1D3"	
	33868..135899	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="BAC TAMU8P4"	
	134292..149379	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1F18"	
	146795..164143	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1K18"	
	162663..193937	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1N23/YUP16B9.IB1O24"	
	183320..206606	
source	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="BAC TAMU2D4"	
	685..1806	
gene	/gene="dl3200c"	
gene	complement(join(695..744,829..903,1053..1232,1804..1806))	
	/gene="dl3200c"	
CDS	complement(join(695..744,829..903,1053..1232,1804..1806))	
	/gene="dl3200c"	
	/note="strong similarity to ribosomal protein L41, Candida maltosa"	
	/codon_start=1	
	/product="ribosomal protein"	
	/protein_id="CAB10211.1"	
	/db_xref="GI:2244789"	
	/translation="MVNPKTKNTYCKNCKEKKHTLHKVTQYKKGKDSLAAGKRRYD RKQSGYGGGTKEVPHKKATTKKIVLRLLQCSCKHSORPIKRCRHFETGGGKGGKGT SLF"	
exon	complement(685..744)	
	/gene="dl3200c"	
intron	number=1	
	complement(745..828)	
	/gene="dl3200c"	
exon	number=1	
	complement(829..903)	
	/gene="dl3200c"	
intron	number=2	
	complement(904..1052)	
	/gene="dl3200c"	
exon	number=2	
	complement(1053..1232)	
	/gene="dl3200c"	
intron	number=3	
	complement(1233..1803)	
	/gene="dl3200c"	
exon	number=3	
	complement(1804..1806)	
	/gene="dl3200c"	
gene	number=4	
	2229..7384	
	/gene="dl3205w"	
CDS	join(2229..2837,2914..3069,3161..3268,3358..3468, 3569..4957,5000..5212,6340..6428,6469..6560,6706..6749, 7316..7384)	
	/gene="dl3205w"	
	/note="similarity to kinesin osm-3, Caenorhabditis elegans contains EST gb:496806"	
	/codon_start=1	
	/product="kinesin like protein"	
	/protein_id="CAB10212.1"	
	/db_xref="GI:2244790"	
	/db_xref="SPTREMBL:O32291"	
	/translation="WAPTPSSRSNQYTLIRPTQKRLNFKHKTNPQSGKDPSP PEHVEVIGIRIDYDRKEKSPSILQVNTDQTVRVADGVYDFLDGVFSQEGE BEFYKPIEERIKGVKGNKCTIMMYGPTGAKSHTFGCGKEGIVVRSRLDILGDS DQGVTFVQTVLVNBEIYDLSTNSNNGIGMPKASTKVLEVMGKKNASF ISSTRAGKISKEIVKVERIVKSTLCHERSRSRHCIIIDVPTVGGRLMLVDAGSE NIDQAGQGFEMKQAKINQGNIALKRVESIANGDSHVPDRSKLTMLODSFDD KSKMLICASPDPKEMHKLCTLEYGAKACIVRGSHIPNCKYGGESASVILGS RIAAMDEFITIKQSEKKQKERNKAKOLKKKEEVAALRSLLTQREACATNEEIK EKNVETQLLSELDKLEECRMAEFVEMRMEERIIVQQBELEMMRRLEIE VEFRRNSGVDETSQFARLRSLYSDDDPGMVKSMIDMGDPFVKQVWGAHVHQS NTISGNFTNLOKPSNNLTOMYPDRVCLSTVPEEVEEKEKVI VEDKSIICLIT PMSINSEGLKGCNFGADKESASRLRIONFTLCQNGRELSSHSGOEEQANI ASPDKQNFSTITNKAALAVEAKENNISVDQRENGQLDIYVKNETAADNPKLIT TLRTVTDATLADRLKLIIEYIGSDNQAFETFLKIGVINLQQAQAFHFYLPFCQAQV AKEKSTVQATSLTERIGADYINLSLHFLILFPFGESHQHRMLSPHSHQTSPLSS VTFNHQILIKTSLERIGADYINLSLHFLILFPFGESHQHRMLSPHSHQTSPLSS FPLLSRADDEPLVDITPLFEVVLGFTLVLELAVVNDRIAG"	
exon	2229..2837	
	/gene="dl3205w"	
intron	number=1	
	2838..2913	
	/gene="dl3205w"	
	number=1	
	2914..3069	
	/gene="dl3205w"	
exon	number=2	
	3070..3160	
	/gene="dl3205w"	
	number=2	
	3161..3268	
	/gene="dl3205w"	

170129 CTGGGATACCAGTGACTAGACTTTGACACAGAACGAGAAGAGGCTAAATTTCTCTTGCTG 170188

1856 ATAGGTTGCATAAGCGGGTTGTGGGACAGAAATCAAGCGGTAAATGACAGTTCTGTAGGCGAA 1915

170189 AATGTTGCAGAGAGTGTGTGGACAGACGAACTGTGAAGCTGTGGCTGCTGCACAA 170248

1916 TTCTAAGGTCAAGCGGACGACTTTGGTAGGGCAACAACGCGCACTGATCATTTCTATTCC 1975

170249 TACTAAGGTCTGAGGGTTGGACTAGGAAGGCGACAACAACCGTCTGGTCTCTCTCTCC 170308

1976 TTGGACCAACTGTGTGTGGCAAACTGAGCTCCCGAAGGCTCTTGTCTGAGCAGCTGTGTTG 2035

170309 TCGETCCAACTGTGTGTGTGTAAACCGAGCTGCTAAGGCTCTGGCTGAGCAACTCTTGG 170368

2036 ATGATGAAAACTCTTAGTTTCGGATGTATGTGCGAAATATATGGAAACAACACTCTGTCT 2095

170369 ACAGTGAAAACCTACTCGTTAGGCTTGATATGTGCGAGTATATGACAAATTTCTCTGTTA 170428

2096 CTCGCCCTCATTTGGGGCACCAAGGGTATG----- 2125

170429 ATAAACTTATAGGGGCAACACTGGATGTGCTTAAACTCAAGCCACTCTTTATGTCT 170488

2126 -----T 2126

170489 TTTTGTGATTAATGGTATAATGGAACTCACATGATACTTGTGTTTTTCTTAGGTACAT 170548

2127 TGGTCACGAGGAGGTGGACAACTTACTAGGCTGTGAGAGCGGACCTTATGTGTCTAT 2186

170549 TGGTCATGAAGAAGGTGGACAACTTACTGAACCTGTGAGGAGGACCTTACTGTGTGT 170608

2187 ACTCTTTGATGAAGTGAGGAAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCCAAGTTTT 2246

170609 ACTATTTGATGAAGTTGAAAGACCCATGTAACTGTATTCAACAGCTACTTCAAGTGCT 170668

2247 GGATGATGTCGATTGACACGCGGCAGGACGACGTCGATTTCAGAACCTCGGTGAT 2306

170669 AGAAGCGGGAGATTAAACCCATTAGCATGGGAGGACGTCGATTTTAAGAACACAGTGAT 170728

2307 AATCAATGACATCAAACTTGCTGCTGAACACCTCTTTCAGGGCTTAACCTGGGAAAGTAAAC 2366

170729 CATATGACITCTTAACCTTGTCAGACACCTTGTTTCHAGGGCTTACAGGTGAATGAC 170788

2367 AATGGAAGTGCCCGGACCTGTGTGATGTCGGGAGGTGAGGAACAATTCAGACACAGAGCT 2426

170789 AATCAAAGTGGCTCGAGACAAACGCAATGAAGAATGCGAAAAACATTTCAGGCGCAGAGCT 170848

2427 CTTCAACAGGCTTGACAGATTTGTGTGTTTCGACCCCTTTACATGACCCAGTTGAGGAA 2486

170849 GTTAAACAGATGGAATGAGATAGTTATGTTTCATCTCTTGCTCCACGAGATTGGCAAA 170908

2487 AGTAGCTCGGCTCAAAATGAAGAGCTGTGTCGCGCTTGTGAAAGAGGAGTTGCTTT 2546

170909 AATTGTTCAATCCAAAGTGAAGAATGTTGTGCT----- 170941

2547 GGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGAGAGTTATGACCCGCTGTATG 2606

170942 -----TGCTTGAGGTTTACATTTTATAATATTGTTATGTTATTTGTAGTGTATG 170996

2607 TGCT--AGGCCATATAGGAGATGGATGGAGAAGGTGGTGACAGAACTGTCAAAAGATG 2664

170997 TGCTAGAGGCCCTATTGAAGAATGGCTAGAGAGGAAAGTGGTCAACAGATATATCAATGATG 171056

2665 GTTGCGGTGAGGAATCGATGAACCTCACTGTTTACATAGATGACGCGCTGGTGAT 2724

171057 ATTGTGCGAGAAGAAATCGGTGACGATTCATATTGTGTGCATTTGATGTAAACGAGGCTAA 171116

2725 CTGTGTACCGGGTAGAAAGTGGAGGCTAGTGGACGCTTCAACAGGCAAGAGTCAGAT 2784

171117 ACTGATCTTGTCTACCGNATAGATAAAAATGTTTTGTGAAGATAGAGCAAACTTAGAT 171176

2785 GTGCTGATTTCATATTCGCTTAACGGGCCAAAGAGAGTGTATG 2824

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

on: February 13, 2004, 01:34:21 ; Search time 535 Seconds
(without alignments)
4596.613 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

uence: 1 MNPEKFTHTNETIATAHEL.....VKKWRIEEIEDDDNEEMIED 911

ring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

arched: 2552756 seqs, 1349719017 residues

al number of hits satisfying chosen parameters: 5105512

imum DB seq length: 0

imum DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

mand line parameters:

DEL=frame+ p2n.model -DEV=xlp
cg2 1/USPTO spool_P/US09812350/runat 10022004_101514 13343/app_query.fasta_1.1095
N_Geneseq 19Jun03 -QWTFastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0
PEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
ST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
DE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
CGN 1 1 396 @runat 10022004_101514 13343 -NCPU=6 -ICPU=3
MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSELOCK=100 -LONGLOG
TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
PEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

abase :

N_Geneseq 19Jun03:.*
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	4550	100.0	3105	22	AAI66065	Arabidopsis thalia
2	4545	99.9	2736	24	ABZ12430	Arabidopsis thalia
3	4545	99.9	2736	25	ABZ42033	Arabidopsis thalia
4	4388	96.4	6376	22	AAI66066	Arabidopsis thalia
5	4003	88.0	3049	22	AAI66067	Glycine max heat s
6	3896	85.6	3052	22	AAI66068	Nicotiana tobacum
7	3883.5	85.4	3084	22	AAI66070	Zea mays heat shoc
8	3822.5	84.0	3058	22	AAI66069	Triticum aestivum
9	3790	83.3	2821	22	AAI66075	Triticum aestivum
10	3657	80.4	3942	22	AAI66073	Phaseolus lunatus
11	3657	80.4	3942	22	AAI66074	Triticum aestivum
12	2502.5	55.0	2045	22	AAI66071	Zea mays 101 kDa h
13	2502.5	55.0	2045	22	AAI66072	Zea mays heat shoc
14	2298	50.5	7726	22	AAI66083	Leishmania donovan
15	2288	50.3	5997	22	AAI66084	Leishmania major h
16	2288	50.3	2619	24	ABS63272	High level promote
17	2245.5	49.4	43980	24	ABQ67192	Listeria innocua c
18	2245.5	49.4	3011208	24	ABQ69245	Listeria innocua D
19	2236.5	49.2	2944528	24	ABA03041	Listeria monocytog
20	2224.5	48.9	2800	22	AAI66064	Trypanosoma brucei
21	2224	48.9	2776	22	AAI66057	Synechococcus CLPB
22	2224	48.9	2776	22	AAI66076	Synechococcus sp h
23	2204	48.4	2586	23	AAS56353	Salmonella typhi D
24	2194	48.2	2574	20	AX07111	Staphylococcus aur
25	2193	48.2	2574	23	AAS52524	E. coli DNA for ce
26	2193	48.2	2574	24	AAD33430	E. coli clpB gene.
27	2185	48.0	34980	24	ABQ81847	Bifidobacterium lo
28	2185	48.0	34980	24	ABQ81848	Bifidobacterium lo
29	2162.5	47.5	2571	23	AAS53400	Haemophilus influe
30	2162.5	47.5	1830121	17	AAT42063	Mycobacterium tube
31	2158.5	47.4	2544	24	ABT05624	Mycobacterium tube
32	2158.5	47.4	4403765	22	AAI99683	Mycobacterium tube
33	2158.5	47.4	4411529	22	AAI99682	Mycobacterium tube
34	2150.5	47.3	3464	22	AAI66062	Corynebacterium gl
35	2150.5	47.3	349980	22	AAH68533	C glutamicum codin
36	2145	47.1	2556	22	AAH68464	C glutamicum codin
37	2141.5	47.1	2607	23	AAS51593	Staphylococcus aur
38	2141.5	47.1	2649	23	AAS54791	Staphylococcus aur
39	2140.5	47.0	23474	23	AAS59511	Propionibacterium
40	2136.5	47.0	6491	20	AAI3093	Enterococcus faeca
41	2136.5	47.0	6491	24	ABS98888	Enterococcus faeca
42	2133	46.9	2565	23	AAS51555	Pseudomonas aerugi
43	2133	46.9	2580	20	AX98121	Nucleotide sequenc
44	2133	46.9	42235	20	AAI98035	Nucleotide sequenc
45	2111.5	46.4	2571	24	ABQ90177	M. capsulatus gene

ALIGNMENTS

RESULT 1

AAI66065

ID AAI66065 standard; DNA; 3105 BP.

AC AAI66065;

DT 14-JAN-2002 (first entry)

XX Arabidopsis thaliana heat shock protein 101 gene (GenBank: U13949).

XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

XX Arabidopsis thaliana; ds.

OS Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAMS1659.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3105 BP; 891 A; 541 C; 852 G; 821 T; 0 other;

ignment Scores:

ed. No.:	0	Length:	3105
ore:	4550.00	Matches:	911
ercent Similarity:	100.00%	Conservative:	0
st Local Similarity:	100.00%	Mismatches:	0
ery Match:	100.00%	Indels:	0
:	22	Gaps:	0

-09-812-350-17 (1-911) x AA166065 (1-3105)

1	MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu	20
163	ATGAATCCAGAGAAATTCACACACAGACAAACAGACAAATTCCTACAGCTCATGAGCTA	222
21	AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer	40
223	GCTGTGAATCAGGACATGCTCAATTCACCTCCCTTTCATTTAGCTGGTCTTGTCTCT	282
41	AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln	60
283	GATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTCCGCTGGCGAGAACGACGCTCAA	342
61	SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro	80
343	CTGCTGAAAGAGTATCAATCAAGCCCTTGAAGAGCTTCCTTCACAACTCTCTCCACCT	402
81	AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln	100
403	GATGATATTCAGCGAGTCTTAGTCTTTATTAAGGTCAATTCGTGCTCAAGCTCTCAG	462
101	LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp	120
463	AAGTCACGAGTGATCTACTATTTGGCTGTTGACCACTGATTTAGTGGCTCTCTTGAAGAT	522
121	SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu	140

Db	523	TCTCAATCAGGATTTGTTGACGAGTCTGTTAGCGAGCGGAGGTAAGTCTTGAG	582
Qy	141	ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn	160
Db	583	TTTGAGAAAGCTTCGTGGGAAAGAGGGAAGAAAGTGTGAGAGTGTCTTCAGGGGACACAAAT	642
Qy	161	PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro	180
Db	643	TTTCAAGCTTTAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAGGAGCTTGATCTCT	702
Qy	181	ValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgThrLys	200
Db	703	GTGATTGCTGCTGATGAGGAGATTAGAAGAGTCTGAGGAGTCTTTCGAGGAGAACGAAG	762
Qy	201	AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu	220
Db	763	AACATCTCTGCTTATTGGAGAGCCAGGAGTGTGTAACACAGCTGTGGTGTGAAGGTTTA	822
Qy	221	AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer	240
Db	823	GCACAAAGGATTGTGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTCG	882
Qy	241	LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluArgLeu	260
Db	883	TTGGACATGGTGGCTTAGTGTGCTGCTAAATACCGAGGAGATTGGAAGAAAGTTG	942
Qy	261	LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu	280
Db	943	AAATCTGTTTGAAGAGATTGAGAGCGCTGAGGCAAGTGATTCCTTTATTTGATGAG	1002
Qy	281	IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe	300
Db	1003	ATTCAATTTGTTCTTGGTCTGCGCAAACTGAAGGGTCGATGGATGCAGCTAACTCTTC	1062
Qy	301	LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlyTyr	320
Db	1063	AAGCCCATGTTAGCTAGAGGCGCTTCGATGCTGCTGCTACACGCTTGAAGAATAC	1122
Qy	321	ArgLysTyrValGluLysAspAlaPheGluArgArgPheGlnGlnValTyrValAla	340
Db	1123	AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTCCACAAAGTCTATGTTGG	1182
Qy	341	GluProSerValProAspThrIleSerIleLeuArgGlyLysGluLysTyrGluGly	360
Db	1183	GAGCCAGTGTGCTGACACCATTAGTATCTTAGAGGACTCAAGAGAGAGTATGAGGA	1242
Qy	361	HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeuSerAlaArg	380
Db	1243	CATCATGTTGCGGAATCCAAAGCAGAGCTCTTATAATGCTGCTCAGCTGTCTGCTCGT	1302
Qy	381	TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla	400
Db	1303	TACATACTGTTGCGCATTTACCGGATAAAGCAATTGATTTGTTGATGAGGCTTGTGG	1362
Qy	401	AsnValArgValGlnLeuAspSerGlnProGluGluIleAspLeuArgLysArg	420
Db	1363	AATGTGAGAGTCCAGCTTGTAGTCAACCTGAAGAGATTGATTAACCTTGAAGGAGAGG	1422
Qy	421	MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla	440
Db	1423	ATGACGCTGGAAATTTGAATTCACGCTTGGAAAGGAGAGGATAAAGCCAGCAAGCT	1482
Qy	441	ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr	460
Db	1483	CGACTATAGAGGTGCGGAAGAGCTTGTGATGACCTGAGAGCAAGCTTTCAGCTCTCACG	1542
Qy	461	MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysArg	480
Db	1543	ATGAATACAGAAAGGAGAGAGAGAAATTTGATGAGATTGGAAGGCTTAAACAGAAAGA	1602
Qy	481	GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAla	500

1603 GAAGAGCTCATGTTTCTTTCTTTCAGGAGGACAGACGAAGATATGACCTTGGCAAGAGCTGCT 1662
 501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
 1663 GATCTAGATATGCGGCAATTCAGAAAGTGAATCTGCAATTTGCCCACTTGAGGAAC 1722
 521 SerSerGluGluAenValMetLeuThrGluAenValGlyProGluHisIleAlaGluVal 540
 1723 TCTTCTGAAGAGAATGTGATGCTCACAGAAACGTTGGCCCTGAACACATTTGCTGAGGTT 1782
 541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAenGluLysGluArgLeu 560
 1783 GTGAGCCGTTGACAGAGGATTCAGTACGAGACTTGGCCAAATGAGAGGAGAGGTTG 1842
 561 IleGlyLeuAlaAspArgLeuHisIleArgValValGlyGlnAenGlnAlaValAspAla 580
 1843 ATTGGTCTTGTGATAGTTGCATAAGCGGTTGTGGGACAGATCAAGCGGTAAATGCA 1902
 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 1903 GTTCTGAGGCAATTCCTAAGGTCAAGGCGAGGACTTGGTAGGGCACAACAGCCAACTGGA 1962
 601 SerPheLeuPheLeuGlyProThrGlyValGlyThrGluLeuAlaLysAlaLeuAla 620
 1963 TCATCTTATCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCCAAAGGCTCTTGCT 2022
 621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGlu 640
 2023 GAGCAGCTGTTTGTATGATGAACCTCTTAGTTCGGATTGATGTCGGATATATGGA 2082
 641 GlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGly 660
 2083 CAACACTCTGCTCTCGCTCATTTGGGCGCACACAGGATGTTGGTTCAGGAGGAGGT 2142
 661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680
 2143 GGACACTCACTAGGCTGTGAGGAGCGACTTATTTGTGCATACCTCTTTGTATGAAGTG 2202
 681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
 2203 GAGAAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAGTTTGGATGATGTCGATG 2262
 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMetThrSerAsn 720
 2263 ACAGCGGCAAGGAGGAGGAGTCAATTCAGAACTCGGTGATAATCATGACATCAAC 2322
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2323 CTGCTGCTGAACACCTCTTGCAGGCTCACTGGAAAGTAACATGGAAAGTGCCCGG 2382
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
 2383 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACAGAGCTTTGAACAGGCTTGAC 2442
 761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
 2443 GAGATTGTTGTTGTCACCCCTTTCATCATGACAGTTGAGGAAGTAGTTCGCTTCAA 2502
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2503 ATGAAGACGTTGCTGCTCGGCTTGTCTGAAGAGGAGTTGCTTTGGCAGTCATGTAGCT 2562
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2563 GCTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGCTGTATGCTGCTAGCCCTATAAGG 2622
 821 ArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
 2623 AGATGGATGGAGAGAGGTTGGTCAACAAGATGTTGTGCGTGAAGAAATC 2682
 841 AspGluAenSerThrValTyrIleAspAlaGlyValAspLeuValTyrArgValGlu 860
 2683 GATGAAGAACTCCACTGTTTACATAGATGACAGGCGCTGGTATCTTTGTGACCGGAGTAGA 2742

QY 861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 Db 2743 AGTGGAGGCTCTAGTGGACCTTCAACAGCAAGAGTCAGATGCTGCTGATTCTATTTGCT 2802
 QY 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
 Db 2803 AACGGGCCAAAGAGAGTGTATGATGCTCAGCGGTGAGAGAGATGAGATCGAGGAATA 2862
 QY 901 GluAspAspAsnGluGluMetIleGluAsp 911
 Db 2863 GAAGATGACGATTAATGAGGAATGATCGAGGAT 2895

RESULT 2

ABZ12430
 ID ABZ12430 standard; DNA; 2736 BP.
 XX
 AC ABZ12430;
 XX
 DT 21-JAN-2003 (first entry)
 XX
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 235.
 XX
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200216655-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26685.
 XX
 PR 24-AUG-2000; 2000US-227866P.
 PR 26-JAN-2001; 2001US-264847P.
 PR 22-JUN-2001; 2001US-300111P.
 XX
 PA (SCRI) SCRIPPS RES INST.
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Harper JF, Kreps J, Wang X, Zhu T;
 XX
 DR WPI; 2002-304127/34.
 XX
 PT Identifying a stress condition to which a plant cell has been exposed
 PT and producing plants with increased tolerance to these abiotic stresses
 PT

PS Claim 144; SEQ ID NO 235; 577pp + Sequence Listing; English.

XX The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising:
 CC (a) contacting nucleic acid representative of expressed polynucleotides
 CC in the plant cell with an array or probes representative of the plant
 CC cell genome; and
 CC (b) detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ12197/4) used
 CC in methods of the invention.

CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.

SQ Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 2736
 Score: 4545.00 Matches: 910
 Percent Similarity: 99.89% Conservative: 0
 Best Local Similarity: 99.89% Mismatches: 1
 Query Match: 99.89% Indels: 0

09-812-350-17 (1-911) x ABZ12430 (1-2736)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGAATCCAGAGAAATTCACACACAGACAAACGAGACAATTTGCTACAGCTCATGAGCTA 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
61 GCTGTGATCCAGACACATGCTCAATTCATCTTTGCAATTTAGCTGGTCTTTGATCTCT 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAsnAlaAlaGln 60
121 GATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTCCGGTGGCGAGAACGAGCTCAA 180
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
181 TCTGCTGAAGAGTGATCAATCAAGCGCTTGAAGAGCTTCTCTTCAATCTCTCTCCACCT 240
81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaGln 100
241 GATGATATTCAGCGAGTCTAGTCTTATTAAGTCAATTCGTCGCTCAAGCTGCTCAG 300
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp 120
301 AAGTCACGAGGTGATATCTATTTGGTGTGTGACCACTGATTATGGGTCTCTTGAAGAT 360
121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
361 TCTCAATCAGGAGTTTGTGAACGAGTGGTGTAGCGACGGCGAGGTTAAGTCTGAG 420
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
421 GTTGAGAGCTTCTGCGGAAAGAGGAGGAGAGTGTGAGAGTCTTCAAGCGACACAAAT 480
161 PheGlnAlaLeuLysThrThrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
481 TTTCAAGCTTTAAGACTTATGGAAGAGATTTGGTTGAGCAAGCGAGGAGCTTATCTCT 540
181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
541 GTGATTTGGTCTGATGAGAGATTTAGAAGAGTCTGAGAGTCTTTCGAGGAGAGAGGAG 600
201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
601 AACAACTCTGTCTTATTCGAGAGCCAGAGTGTGTAACAGCTGTGTTGAAGGTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGGATTTGTGAAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTGC 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGluGluArgLeu 260
721 TTGGACATGGTGGCTTAGTGTCTGTGTCTTAATACCGAGGAGAGGTTTGAAGAAAGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
781 AAATCTGTTTGAAGAGTTCAGGACGCTGAGGCGCTGAAGCAAGTCAATCTCTTTATTCAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTTGGTCTTCTGGTGTGCAAAATCTGAAGGGTGCATGATGACAGCTAATCTGTT 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyValAlaThrThrLeuGluGly 320
901 AAGCCCATGTTAGTAGGGGAGCTTTCATGCTATGTTGCTTACCAACGCTTGAAGATAC 960
321 ArgLysThrValGluLysAspAlaAlaPheGluArgArgPheGlnValThrValAla 340
961 AGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTCCAAACAGTCTATGTTGG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysThrGluGly 360

1021 GAGCCAAGTGTGCTGACACCATTAGTATCTTAGAGGACTCAAGAGGAGATATGACGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGTTGTGCGAATCCACAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTCGT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATACTGCTGGCATTACCGGATTAAGCAATTCATTTGTTGATGAGGCTTGTGCG 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
1201 AATGTGAGAGTCCAGCTTGTAGTCAACCTGAAGAGATTGATAACCTTGAAGAGGAGG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
1261 ATGCACTGGGAAATTTGAACCTTCAGCGCTTGAAGAGGAGGATTAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGTTCGCGAAGAGCTTGTAGTACCTGAGACAGAGCTTCAGCTCTCAG 1380
461 MetLysThrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
1381 ATGAATACAGAAAGGAGAAAGAGAGATTCATGAGATTCGAAGGCTTAAACAGAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgGlyrAspLeuAlaArgAlaAla 500
1441 GAAGAGCTCATGTTTTCTTTCAGAGGAGGAGAGAGATGATGACCTTGAAGAGCTGCT 1500
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1501 GATCTAAGATATGCGCAATTCAGAGTGAATCTGCAATTCGCCAATCTTGAAGGAACT 1560
521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAAGAGATGTGATGCTCACAGAAAGCTTGGGCGCTGAACACATTTGCTGAGGT 1620
541 ValSerArgTyrThrGlyIleProValThrArgGluGlyGlnAsnGluLysGluArgLeu 560
1621 GTGAGCGCTTGACAGGGATTCAGTACGAGACTTGGCCAAATGAGAGGAGAGGTTG 1680
561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
1681 ATTGCTCTGCTGATAGTGTTCATAGCGGTTGTGGACAGAGTCAAGCGTAAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
1741 GTTCTGAGGCAATTTCAAGGTCAAGGCGAGGACTTGGAGGCGCACACAGCCAACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATTTCTTATTCCTTGACCCACTGCTGTGGCAAACTGAGCTGCCAAGGCTCTTGTCT 1860
621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluThrMetGlu 640
1861 GAGCAGCTGTTGATGATGAAAACCTCTTAGTTCGATGATGATGTCGGAATATATGGA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 560
1921 CAACACTCTCTCTGCTGCTCAATGGGGCACACCAGGGGTATGTTGTCACAGGAGGT 1980
661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680
1981 GGACAACTAATGAGGCTGTGAGGAGGCGACCTTATTTGTCATACTCTTTGATGAAGTG 2040
681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
2041 GAGAAAGCTCATGTTGCTGCTTCAACACTCTGCTCAAGTCTTGGATGATGTCGATG 2100
701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720

2101 ACAGCGGCGAGCAGCAGCAGTCGATTTCAGGAACCTCGGTGATATCATGACATCAAC 2160
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2161 CTTGGTCTGACACCTCTCTTGAGGGCTAATGGGAATACATGGAAGTGGCCGG 2220
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAenArgLeuAsp 760
 2221 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACCAGAGCTCTGAACAGGCTTGAC 2280
 761 GluLeuValValPheAspProLeuSerHisAspClnLeuArgLysValAlaArgLeuGln 780
 2281 GAGATTGTGTGTCGACCCCTTCATGACACAGTGTGAGGAAGTAGCTCGCTCAA 2340
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2341 ATGAAGACGCTGCTGTCGGCTTGCTGAAAGAGGAGTTCCTTTGGCAGTCACTGATCT 2400
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2401 GCTTTGACTATATCTTTGGCAGAGATTATGACCCGGTGTATGCTGCTAGGCCCTATAAGG 2460
 821 ArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
 2461 AGATGATGAGGAAGAGGTGTCAGAGAACTGTCAAGAGTGTGTGCGTAGGAATC 2520
 841 AspGluAenSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
 2521 GATGAAGAACTCCACTGTTATAGATGACAGGCGCTGCTGATCTTGTGTACCGGGTAGAA 2580
 861 SerGlyGlyLeuValAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 2581 AGTGGAGGTCTAGTGGACGCTTCAACAGGCAAGATCAGATGTGCTGATTATTTGCT 2640
 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
 2641 AACGGCCAAAGCAAGTATGACGCTCAGCGGTGAGAGATGAGATCGAGGAATA 2700
 901 GluAspAspAspGluGluMetIleGluAsp 911
 2701 GAAGATCAGCAATATGAGGAATGATCGAGAT 2733

JLT 3

ABZ42033

ABZ42033 standard; cDNA; 2736 BP.

ABZ42033;

27-FEB-2003 (first entry)

Arabidopsis thaliana gene #17 modulated by PTGS.

Posttranscriptional gene silencing; PTGS; plant; transformation; gene; ss.

Arabidopsis thaliana.

Key Location/Qualifiers
 CDS 1..2736
 /*tag= a

W0200281695-A2.

17-OCT-2002.

05-APR-2002; 2002WO-EP03806.

06-APR-2001; 2001US-282049P.

(SYGN) SYNGENTA PARTICIPATIONS AG.
 (FRIE-) FRIEDRICH MIESCHER INST.

Zhu T, Glazov EA, Meins F, Wang X, Chang H;

XX WPI; 2003-103337/09.
 DR P-PSDB; ABP81189.
 XX
 PT Novel polynucleic acid segment useful for modulating gene expression
 PT within a cell by posttranscriptional gene silencing, and for augmenting
 PT a plant cell genome -
 XX
 PS Claim 18; Page 185-186; 438pp; English.
 XX
 CC The invention relates to a novel isolated polynucleic acid segment
 CC modulated within a cell by posttranscriptional gene silencing (PTGS). The
 CC invention specifically relates to a method to identify an expression
 CC product that is modulated by PTGS. The polynucleotide is useful for
 CC modulating the gene expression within a cell by PTGS, by introducing the
 CC polynucleic acid into a cell and expressing the nucleic acid segment in
 CC the cell to form a product. The polynucleic acid segment is also useful
 CC for augmenting a cell genome, and for augmenting a plant genome, by
 CC contacting a plant cell with the segment to produce a transformed plant
 CC cell, and growing the transformed plant cell to produce a differentiated
 CC transformed plant. The sequences shown in ABZ42017 - ABZ42142 represent
 CC segments of A. thaliana cDNA modulated by PTGS.

XX Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	2736
Score:	4545.00	Matches:	910
Percent Similarity:	99.89%	Conservative:	0
Best Local Similarity:	99.89%	Mismatches:	1
Query Match:	99.89%	Indels:	0
DB:	25	Gaps:	0

US-09-812-350-17 (1-911) x ABZ42033 (1-2736)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 DB 1 ATGAATCCAGAGAAATTCACACAAAGCAACATTCCTACAGCTCATGCTCATGAGCTA 60
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 DB 61 GCTGTGATGTCAGGACATGCTCAATTCCTCTCTTGGATTAGCTGTGCTTGTATCTCT 120
 QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 DB 121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGGAGAACGACGCTCAA 180
 QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 DB 181 TCTGCTGAAGAGTGTATCAATCAAGCCTTGAAGAGCTTCCTTCACATCTCTCCACCT 240
 QY 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAlaGln 100
 DB 241 GATGATATTCAGCGCTTCTAGCTTATTAAGTCAATTCCTGCTCAAGCTGCTCAG 300
 QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
 DB 301 AAGTCACGAGGTGATCTCAATTTGGCTGTTCACCACTGATATATGGGTCTCTTGAAGAT 360
 QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
 DB 361 TCTCAATCAGGGATTTTGTGAACGAAGTCTGGTGTAGCGAGCGGGGTAAAGTCTGAG 420
 QY 141 ValGluLysLeuArgGlyLysGluLysLysValGluSerAlaSerGlyAspThrAsn 160
 DB 421 GTTGAGAGCTTCTGGGAAGAAGAGGAGAAAGTTGAGAGTCTTTCAGGGGACACAAAT 480
 QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
 DB 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGAGGAGCTTGATCT 540
 QY 181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200

541 GTGATTGGTCGTGATGAGGAGATTAGAAGAGTCGTGAGGATTCTTTTCGAGGAGAACGAAG 600
201 AenAnProValLeuIleGlyGluProGlyValGlyLysThrAlaValAlaValGluGlyLeu 220
601 AACATCTCTGCTTATTGGAGAGCCAGAGTGTGTAAACAGCTGTGTGTAAGGTTTA 660
221 AlaGlnArgIleValLysGlyAspValProAenSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGGATTGTAAAGAGAGATGTGCCAACAGCTCTTACTGATGTGAGATTAAATTCG 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGluGluArgLeu 260
721 TTGGACATGGGTGGTGTAGTCTGTGTCTTAATACCCGAGGAGAGATTGCAAGAAAGGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIlePheIleAspGlu 280
781 AAATCTGTTTGAAGAGAGTTCAGGAGCGCTGAAGCGCAAGTGAATCTCTTATTGATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAenLeuPhe 300
841 ATTCAATTTGGTCTGTGCTGCGCAAACTGAAGGCTCGATGATGCGAGCTAAATCTGTTTC 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrIleuGluGluTyr 320
901 AAGCCCATGTTAGTAGGGGAGCTTCGATGCATTTGGTGTACAACTTGAAGGATAC 960
321 ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrValAla 340
961 AGCAATAATGTTGAGAAAGATGCTGCTTTTGAGAGAGAGGTTCCAAACAGTCTATGTTGG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
1021 GAGCCAAAGTGTCCCTGACACCAATTAGTATCTTTAGAGGACTCAAGGAGAGATGAGGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAenAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGTTGCGAATCCAGACAGAGCTCTTATAATGCTGCTCAGCTGCTCTCTGT 1140
381 TyrIleThrGlyArgHisLeuProLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATAACTGGTTCGGCATTTACCGGATAAAGCAATTGATTTGGTTGATGAGGCTTGTGCG 1200
401 AenValArgValGlnLeuAspSerGlnProGluGluIleAspAenLeuGluArgLysArg 420
1201 AATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAACCTTTGAAGAGAGG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAspLysAlaSerLysAla 440
1261 ATGCAGCTGGAATTTGAATTTCAAGCTTTGAAAGGAGAGATTAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGTGGGAAAGAGCTTGATGACCTTGATGACCTGAGAGACAAGCTTCTCAG 1380
461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
1381 ATCAATAACAGAAAGAGAGAGAGATTTGATGAGATTGCAAGGCTTTAAACAGAAAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
1441 GAAGAGCTCATGTTTCTTTCAGGAGGCGAGAACGAGATATGACCTTGCAGAGCTGCT 1500
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1501 GATCTAAGATATGGCGCAATCAAGAAGTGGAAATCTGCAATTTGCCCAACTTGAAGAACT 1560
521 SerSerGluGluAenValMetLeuThrGluAenValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAAGAGAAATGTGATGCTCAAGAAAAAGTTGGGCCCTGGAACACATCTGCTGAGGTT 1620
541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAenGluLysGluArgLeu 560
1621 GTGAGCGGTTGGACAGGATTCAGTGAAGAGACTTGGCCAAATGAGAGAGAGGTTG 1680

561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAenGlnAlaValAenAla 580
1681 ATTGGTCTCTGATAGTGTGCATAGCGGTTGTGGACAGAAATCAAGCGGTAAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
1741 GTTTCTGAGCAATTTCTAAGTCAAGGGCAGGACTTGGAAAGGCCACACAGCCAACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATCTTATTTCTTGGACCACTGTTGTGGCAAACTGAGCTCGCCACAGGCTCTTGCT 1860
621 GluGlnLeuPheAspAspGluAenLeuValArgIleAspMetSerGluLysMetGlu 640
1861 GAGCAGCTGTTTGTATGATGAAACCTCTTTAGTTCGGATTGATGATGCGAAATATATGGA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 660
1921 CAACACTCTGCTCTCGCTCATTTGGGGCACCACCGAGGTATGTTGGTCACGAGGAAGT 1980
661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680
1981 GGAACAATACTGAGGCTGTGAGAGGGGACCTTATTTGTCTACTCTCTTTGATGAAGTG 2040
681 GluLysAlaHisValAlaValPheAenThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
2041 GAGAAAGCTCATGTTGCTGCTCTCAACACTCTCTGCTCAAGTTTGGATGATGCTCGATTG 2100
701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
2101 ACAGACGGGCAAGGAGGAGCAGTCGATTTTCAGAACTCGGTGATATCATGACATCAAC 2160
721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
2161 CTTGGTGTGAAACACTCTCTTGCAGGGCTAACTGGGAAAGTAACAATGGAAGTGGCCGG 2220
741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuAenArgLeuAsp 760
2221 GACTGTGTGATGGGGAGGTGAGGAAACACTTCAGACCAAGAGCTCTTGAACAGGCTTGAC 2280
761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
2281 GAGATTGTGTGTTCAGCCCTTTTACATGACACAGTCAGTTGAGAAAGTAGCTCGGCTTCAA 2340
781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
2341 ATGAAAGACGTTGCTGCTCCGCTTGTCTGAAAGAGAGTTCCTTTGGCAGTCACTGATGCT 2400
801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
2401 GCTTTGAGCTATATCTTTGGCAGAGATTATGACCCGGTGTATGCTGTAGGCTTAAAG 2460
821 ArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
2461 AGATGATGAGAGAGAGAGTGTGACAGAACTCTCAAGATGTTGTGTGTGAGGAAATC 2520
841 AspGluAenSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
2521 GATGAAACCTCCACTGTTTACATAGATGACGGCGCTGGTGTATCTTGTGTACCGGCTAGAA 2580
861 SerGlyGlyLeuValAspAlaSerThrGlyLysSerAspValLeuIleHisIleAla 880
2581 AGTGGAGGCTTAGTGGACGCTTCAACAGGCAAGAGTCAGATGCTGATGCTCATATTGCT 2640
881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMetArgIleGluGluIle 900
2641 AACGGGCAAGAGAGAGTGTGATGCTCAGCGGTGAGAGATGAGATCGAGGAAATA 2700
901 GluAspAspAsnGluGluMetIleGluAsp 911
2701 GAAGATGACCAATAATGAGGAAATGATCGAGAT 2733

ULT 4

66066

AAI66066 standard; DNA; 6376 BP.

AAI66066;

14-JAN-2002 (first entry)

Arabidopsis thaliana heat shock protein 101 gene (GenBank: AF218796).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

Arabidopsis thaliana; ds.

Arabidopsis thaliana.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51659.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 6376 BP; 1914 A; 1150 C; 1470 G; 1842 T; 0 other;

gment Scores:

d. No.:	1.31e-315	Length:	6376
Matches:	4388.00	Matches:	908
Conservative:	86.56%	Conservative:	0
Local Similarity:	86.56%	Mismatches:	3
Indels:	96.44%	Indels:	140
Match:	22	Gaps:	4

09-812-350-17 (1-911) x AAI66066 (1-6376)

1 MetAnProGluLysPheThrHisLysThrAnGluThrIleAlaThrAlaHisGluLeu 20

1132 ATGAATCCAGAGAAATTCACACACAGACAAACAGACAAATGTACAGCTCATGAGCTA 1191

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40

1192 GCTGTGAATGAGGACATGCTCAATTCCTCTTGCAATTTAGCTGGTCTTGTCTCTCT 1251

Qy	41	AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAsnAlaGln 60
Db	1252	GATCCACCGGTATTTCTCAAGCAATCTTAGTCCCGTGGGAGAACGAGCTCAA 1311
Qy	61	SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
Db	1312	TCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAGCTTCTTCACAATCTCTCCACCT 1371
Qy	81	AspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAlaGln 100
Db	1372	GATGATATTCACGAGATTTCTAGTCTTATTAAAGTCAATTCGTCTCAAGCTGTCTAG 1431
Qy	101	LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
Db	1432	AACTCAGAGTGATCTACTCATTTGGCTTTGACCACTTGATTTATGGTCTTCTTGAGAT 1491
Qy	121	SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
Db	1492	TCTCAATCAGGATTTGTTGAACGAAGTCGGTGTAGCGACGCGAGGGTAAAGTCTGAG 1551
Qy	141	ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
Db	1552	GTTGAGAGCTTCGTGGAAAGAGGAAAGAGTTGAGAGTGCTTTCAGGGGACACAAAT 1611
Qy	161	PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
Db	1612	TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGACGACGAGGAGCTTGATCTCT 1671
Qy	181	ValIleGlyArgAspGluLurIleArgArgValValArgIleLeuSerArgThrLys 200
Db	1672	GTGATTCGTCTGATGAGGAGATTAGAGAGTCGTGAGGATCTTTCGAGGAGAGACGAAG 1731
Qy	201	AsnAspProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
Db	1732	AACATCTCTGCTGCTTATGAGAGAGCAGAGTTGGTAAACAGCTGTGTTGAAGGTTTA 1791
Qy	221	AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuLeuSer 240
Db	1792	GCACAAAGGATTTGTAAGAGAGATGTGCCACACAGCTTACTGATCTGAGATTAATTCG 1851
Qy	241	LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu 260
Db	1852	TTGGACATGGTGTCGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1911
Qy	261	LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
Db	1912	AAATCTGTTTGAAGAGAGTTGAGGACGCTGAAGGCAAGTGAATCTCTTTTATGATGAG 1971
Qy	281	IleHisLeuValLeuGlyAlaGlyLysThrGlySerMetAspAlaAlaAsnLeuPhe 300
Db	1972	ATTCAATTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2031
Qy	301	LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
Db	2032	AAGCCCATCTGTTAGCTAGAGGCGAGCTTCGATTCATTTGTTGCTACACGCTTGAAGATAC 2091
Qy	321	ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrValAla 340
Db	2092	AGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTTCAACAGCTTATGTTGG 2151
Qy	341	GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyrGluGly 360
Db	2152	GAGCCAAAGTGTGCTGACACCATTAGTATCTCTTAGAGGACTCAAGGAGAGTATGAGGGA 2211
Qy	361	HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
Db	2212	CATCATGTTGCTGCAATCCACAGACAGCTCTTATAATGCTGCTGCTGCTGCTGCTGCTGCT 2271
Qy	381	TyrIleThr----- 383
Db	2272	TACATTAACCTGGTATGTTAAGATCTTAATCTTAAGCTGATGTTTATGGTTTTCATATAGTG 2331
Qy	384	-----GlyArgHisLeuProAspLysAlaIle 392

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51662.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid family amino acid sequence (AA166084) encoding a plant Heat Shock Protein (HSP) increased stress tolerance, especially to heat. The transgenic plant has grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3049 BP; 850 A; 564 C; 863 G; 772 T; 0 other;

ignment Scores:

d. NO.:	1.94e-287	Length:	3049
re:	4003.00	Matches:	794
cent Similarity:	94.85%	Conservative:	72
t Local Similarity:	86.97%	Mismatches:	39
ry Match:	87.98%	Indels:	8
	22	Gaps:	6

09-812-350-17 (1-911) x AA166067 (1-3049)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
136 ATGAATCTGAGAGTTTACTCACAAGACTAATGAAGCTTGTGGTGGCCGAGCTC 195
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisIleuAlaGlyAlaLeuIleSer 40
196 GCGATGAGTTCAGCCACGCGCAATTGACTCCATCCACTTGGCCCAACGCGCTGATTTCC 255
41 AspProThrGlyIlePheProGlnAlaIleSerAla---GlyGlyGluAsnAlaAla 59
256 GATCCCAACGCACTTCGTGCTAGCAGATACAGCGCGCGCGGAGGAGTTCGCA 315
60 GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 79
316 CGCGCGGAGGAGGAGTTCGCAACGAGCTCTGAAGAAGCTACCTCCAGTCCCTCCG 375
80 ProAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAla 99
376 CCGGACGAGTGCAGCGGAGCAGCAACCTCTGAGGGCCATCAGGAGGACACAGCGCGC 435
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
436 CAAATATCAGTGGCGACACCGCTTGGCCCGTTGATCAGTTGATCTCGGAATCTCTCGAA 495

QY 120 AspSerGlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
Db 496 GACTCCCAATCGAGACCTGTTGAAGGAGCGGGGTTGCGGTGGCGAGGTAGATCG 555
QY 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
Db 556 GAAGTGGATAAGCTTCGTGGGAAGGAGGAGGAGGTTGAGAGCGCTTCCGGGGATACG 615
QY 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAsp 179
Db 616 AATTTCCAAAGCTTTGAAGACTTATGCGCGCTGACCTTGTGTGAACAGCGGGAAGCTCGAC 675
QY 180 ProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThr 199
Db 676 CCGTGTATTTGGCCGTCAGCAGAGATTAGAGGGTTGTAGGATTTCTATCAGGAGGACT 735
QY 200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
Db 736 AAGAACACACCCGCTTCTCGTTGGAGAACCGGCTGTGGGAAAACTGCGGTTGTGGAAGCG 795
QY 220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
Db 796 TTGGCACAGAGATAGTAAGAGCGGATGTTCCAAACCAACCTTGTGCTGATGTGAGGCTTAT 855
QY 240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArg 259
Db 856 GCGTGTGATATGGGGCGTGTGTCGCGGGTCCAGTATAGAGGCGAGTTTTCAGAGGCGG 915
QY 260 LeuLysSerValLeuLysGluValGluAspAlaGlyLysValIleLeuPheIleAsp 279
Db 916 TTAAGGCTGTTTGAAGAGCTGAGGAGGCTGAGGAGGAGTATGATCTTCTTATTCAT 975
QY 280 GluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu 299
Db 976 GAGATTCATTTGGTCTCTGCTGCTGAGTAAAGGCTCCATGGATGCTGCTATATCTA 1035
QY 300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyValaThrThrLeuGluGlu 319
Db 1036 TTCAAACCTATGCTTGTGCGCGCCAGCTTAGTGTGATTTGGTCCACACGCTTGAGGAG 1095
QY 320 TyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrVal 339
Db 1096 TACAGGAAGTATGTGAGAGAGGATGCTGCAATTCGAGAGGAGGTTCCAAACAGGTTTGTG 1155
QY 340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLysTyrGlu 359
Db 1156 GCGGAACCTAGTGTGGTGTGATACCATAGCTTCTCTGTTGGCTTGAAGAGAGATATCAA 1215
QY 360 GlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAla 379
Db 1216 GGCCATCAGCGTGTAGAAATTCAGAGCGCTGCTTGGTATGCGCAGCTCAATTGCTTAAC 1275
QY 380 ArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
Db 1276 CGGTATATAACTGGGCGCTCATCTTCTGACAGGCAATGATTTGGTTGACAGGCTTGT 1335
QY 400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
Db 1336 GCAATATTTAGGGTTCAACTTGTATGTCAGCTGAGGAATTTGATACTTGAAGAGGAG 1395
QY 420 ArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLys 439
Db 1396 AGAATGCAGCTAGAGTGGAACTTCATCTCTGGAGAAAGAGAGAGCAAGAGCTAGCAA 1455
QY 440 AlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGlnProLeu 459
Db 1456 GCGCGCTCTGTTGAAGTGGGAAGAACTTGTATGACTTGGAGGACAAAGCTTTCAGCTTGT 1515
QY 460 ThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLys 479
Db 1516 ATGATGAAGTACCGAAAGAGAGAGAGGGTGTGATGAGTTTCGAAGGCTTCAGAAGAAA 1575

480 ArgGluGluLeuMetPheSerLeuGlnAlaGluArgTyrAspLeuAlaArgAla 499
 1576 AGAAGAGAGCTTCTTTTGTCTTACAGAGCTGAGAGAGATATGATCTGGCTAGCT 1635
 500 AlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
 1636 GCACACCTGCATATGAGCAATTCAGAGGTGGAACTCAATACAACTTGAAGGG 1695
 520 ThrSerSerGluGluAanValMetLeuThrGluAanValGlyProGluHisIleAlaGlu 539
 1696 ---AGCACTGAAGAGAACTGATGTTGACTGAACTGTTGACCGGAGCAATAGCTGAG 1752
 540 ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAenGluLysGluArg 559
 1753 GTTGTGAGCGCTGACCGGTATACCACTTACAGGCTTGGCCAAATGAAAGAAAGA 1812
 560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAenGlnAlaValAan 579
 1813 TTGATTTGGACTTGTGTGACAGATTGCAAGCAGAGTTGTAGCAAGCAAGCTGTTAAT 1872
 580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnProThr 599
 1873 GCTGTTCTGAGGCTGTGTTGAGATCTAGAGCTGGGCTAGAGAGCTCAACCAACCACT 1932
 600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
 1933 GGTTCATTCCTATTCCTTGTCTCAACTGGTGTGGTAACTGAGCTTGCMAAGGCTCTT 1992
 620 AlaGluGlnLeuPheAspGluAanLeuLeuValArgIleAspMetSerGluTyrMet 639
 1993 GCTGAGCAACTATTGTATATGAAACCACTGCTGAGAAATGATATGTCGGAATACATG 2052
 640 GluGlnHisSerValSerArgLeuIleGlyValaProGlyTyrValGlyHisGluGlu 659
 2053 GACAGCACTCGTCTTCAAGATTGATGTGTGACCAACAGGATGTTGACATGAGAA 2112
 660 GlyGlyGlnLeuThrGluAlaValArgArgProGlyCysValIleLeuPheAspGlu 679
 2113 GGTGGCAACTAATCAAGCGTGAAGCGAAGACCTTACAGTGTGTGTCTTTTGATGA 2172
 680 ValGluLysAlaHisValAlaValPheAenThrLeuGlnValLeuAspAspGlyArg 699
 2173 GTGAAAGCAATACATCTGTGTTCAATCTCTTCTTCAAGTTTGGACGATGGAGG 2232
 700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAanSerValIleIleMetThrSer 719
 2233 TTAACATGATGACAGAGCGGTACTGTGGACTTTAGAAACACTGATTAATATCACTCA 2292
 720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
 2293 AATCTTGGAGCAGACACCTCTCTCAGTGGACTTTCAGGAAATGTACCATGCAAGTTGCT 2352
 740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAanArgLeu 759
 2353 CGGATCGGTAAATGAGAGAGGTGAGAGGCAATTTAGGCGCAATTTGCTGATGACATT 2412
 760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
 2413 GATGAATTTGCTGATTTGACCCCTTTCACATGACCAATTTGGAGAGGTTCAGAGCTA 2472
 780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
 2473 CAATGAAGATCTAGTAGTCTCTTGTGAGAAAGGAATTCCTTGGCAGTCACTGAT 2532
 800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProfile 819
 2533 GCAGACTAGCATATATACCTTCCGAGAGCTATGATCCCGTGTATGCTGATACCAATA 2592
 820 ArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
 2593 AGGAGTGGCTTGAGAGAGAGGTGTGACAGAGTTCTTAGAATGCTTGTAGAGAGGAG 2652
 840 IleAspGluAanSerThrValTyrIleAspAlaGly-----AlaGlyAspLeuValTyr 857

Db 2653 ATTGATGAGAAATTCACCGTTTACATGATGCTGACCAATGGGGCGAGTTGCTCTAC 2712
 QY 858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeu 876
 Db 2713 CGTGAGAAAAGAAATGAGAGGTTGTTAATCTTACAACTGGGCGAGAGTCCGATATCTTG 2772
 QY 877 IleHisIleAlaAenGly-----ProLysArgSerAspAlaAlaGlnAlaValLysLys 894
 Db 2773 ATTCAATACCTTAATGACATGCACCTAAA---ACTGATGCTGTTCAAGCAGTCAAGAAG 2829
 QY 895 MetArgIleGluGluIleGluAspAspAanGluGlu 907
 Db 2830 ATGAGATTGAGAAATTCATGATGATGAATGGAAGAG 2868
 AC AAI66068;
 XX 14-JAN-2002 (first entry)
 DE Nicotiana tobacum 101 kDa heat shock protein gene (GenBank: AF083343).
 XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana; ds.
 XX Nicotiana tobacum.
 XX WO200170929-A2.
 PD 27-SEP-2001.
 XX 20-MAR-2001; 2001WO-US08836.
 XX 20-MAR-2000; 2000US-190769P.
 PR 18-APR-2000; 2000US-198116P.
 XX (ARCH-) ARCH DEV CORP.
 XX Lindquist S, Queitsch C, Vierling E;
 WPI; 2001-639123/73.
 P-PSDB; AAM51665.
 PT Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -
 XX Claim 4; Page -; 91pp; English.
 CC The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 CC Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.
 XX SQ Sequence 3052 BP; 898 A; 536 C; 836 G; 782 T; 0 other;
 Alignment Scores: 1.69e-279 Length: 3052
 Pred. No.: 3896.00 Matches: 774
 Score:

Percent Similarity: 93.13% Conservative: 75
 t Local Similarity: 84.96% Mismatches: 56
 ry Match: 85.63% Indels: 6
 Gaps: 5

09-812-350-17 (1-911) x AA166068 (1-3052)

1 MetAsnProGluValPheThrHisLysThrAnGluThrIleAlaThAlaHisGluLeu 20
 101 ATGAATCTTGAAAAATTCACCACAGACTACAGAGGCCCTTGCTGGGGCACTCGAGCTA 160
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 161 GCATATCCGACGAGGCATCTCAATTTAGCCTCTGCATATGCTGTGCTTAAATCT 220
 41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGly---GluAsnAlaAla 59
 221 GATCAATGTTGTTATTTTCGACAGCGATTTGTCATGCTGCTGGTGAATGAAGTAGCT 280
 60 GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 79
 281 RATTGCTGGAGCGGTATGATCAAGCGATGAAGAAGCTACCTTCTCAACACCGGCT 340
 80 ProAspAlleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAla 99
 341 CCTGACGAATCCCACTAGCACTTCCCTTATCAAGGTGTTACGCCGACCAATCGTGG 400
 100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuGlu 119
 401 CAGAACTCTCGTGGTGCAGCCATTTAGCAGTGGATCAGTTGATTTTAGGACTGCTAGAA 460
 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThAlaArgValLysSer 139
 461 GACTCCCAATTTGAGATCTTTGAAAGAGCTGGAGTGAGTGATCAAGAGTGAATCA 520
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 521 GAGGTAGAGAACTTAGAGGAAGGAAGGAAGAAAGTCCGAAGTGTCTCAGGGGACACC 580
 160 AsnPheGlnAlaLeuLysThrTyroGlyArgAspLeuValGluGlnAlaGlyLysLeuAsp 179
 581 ACATTCCCAAGCACTCAACACTTATGGCCGCTGATCTTGTGGAACAGCAGGAAAGCTTGAT 640
 180 ProValIleGlyArgAspGluGluIleArgArgValValAlaGlyLeuSerArgArgThr 199
 641 CCCGTGATTGAGGATGAGAAATTAAGAGATCGTTCCGATTTTATCAAGAGGACT 700
 200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValAlaGluGly 219
 701 AAGAACACACCCGGTCTTATTGGAGAGCCCGCTGTGGTAAACAGCAGTGTGTTGAAGGG 760
 220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
 761 CTAGACAGAGATGTGACGTGGGTGATGTTCCAAAGTAAATTTAGCTGATGTTAGGCTTATA 820
 240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyroGlyGluPheGluGluArg 259
 821 GCATTGATATGGAGCGGTAGTTGCTGGAGCTAAGTACAGAGTGAATTTGAAGAGAGG 880
 260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp 279
 881 CTGAAGGCTGTGCTGAAAGAAATTTGAAGAGCGGAAGGAAAGTAAATCTTTCAATTGAC 940
 280 GluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu 299
 941 GAGATACATTATTAGTCTCGGTGGTGGTGGACAGAAGGGTCTATGGATGCTGCTAATCTG 1000
 300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlu 319
 1001 TTTAAGCAATGCTAGCAGAGGTCATTTACGTTGCAATTTGTCACACTACACTCGAGGAG 1060
 320 TyrArgLysTyroValGluLysAspAlaPheGluArgPheGlnGlnValVal 339

Db 1061 TACAGGAAGTATGTTGAGAAGGANGCTGCATTTTGAGAGGCGTTTCCAGCAGGTGTATGTT 1120
 Qy 340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyroGlu 359
 Db 1121 GCTGAGCCTAGTGTGCTGACACTATTAGTATTCTCCGTGGTGGTGAAGGAGGTATGAA 1180
 Qy 360 GlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAla 379
 Db 1181 GGGCATCATGCTGTCAAAATTCAGGACAGAGCTCTTGTAGTGGCTGCCAGCTCTCATCT 1240
 Qy 380 ArgTyroIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
 Db 1241 CGGTACATTACAGCTGCACATCTGCCAGATAGAGCTATTGACCTAGTTGATGAAGCTTGT 1300
 Qy 400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
 Db 1301 GCAAAATGTTAGAGTTGAGCTTGCATCTCAACCTGAGGAAATTCACAACTCTTGAGAGGAG 1360
 Qy 420 ArgMetGlnLeuIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLys 439
 Db 1361 AGAATTCAGCTAGAGTTGAATCTCACTCTCGAAGGAAAGAACACAAAGCTAGCAAA 1420
 Qy 440 AlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeu 459
 Db 1421 GCAGCTCTCATAGAGTGAGGAAGAACTTGATGATTTGAGAGCAAACTCCAACTTTG 1480
 Qy 460 ThrMetLysTyroArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLys 479
 Db 1481 ATGATGAGGTACAAAGAAAGAAAGAAAGATAGATGAGCTCGCAGGCTCAAGCAAAAG 1540
 Qy 480 ArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyroAspLeuAlaArgAla 499
 Db 1541 CGCGATGAGCTCATCTATCTTATCAAGAAAGCTGAAGAGGATATGATCTGGCAGGGCA 1600
 Qy 500 AlaAspLeuArgTyroGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
 Db 1601 GCAGATCTGAGATATGGGCAATTCAGAAAGTGGAAATGCAATAGCAATCTTGAGAGT 1660
 Qy 520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
 Db 1661 ACCTCAGCT---GAAAGTACAATGCTTAACAGAGACTGTGGGTCTCTGATCAGATCGCGAA 1717
 Qy 540 ValValSerArgTyroThrGlyLeuProValThrArgLeuGlyGlnAsnGluLysGluArg 559
 Db 1718 GTTGTGAGTCGTGAGACTGTTATTCGGTCTCAAGGCTTGGTTCAGAAAGAGAGAGAA 1777
 Qy 560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
 Db 1778 CTGATTGGTCTTGGCGATAGATTGCCAACAGAGTGGTGGGCAAGATCATCGAGTTAGA 1837
 Qy 580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
 Db 1838 GCTGTTGCTGAAGCCGCTGTTAAGGTCAGAGCTGGTTAGGAAGGCCACAGCAACCACT 1897
 Qy 600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
 Db 1898 GGTTCATCTCTTTCTTGGGGCCAACTGGTGTGGAAGAGACAGAGCTCGCTAAAGCTCTT 1957
 Qy 620 AlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyroMet 639
 Db 1958 GCAGAGCAGCTCTTGTGATGATGATAAATCTGATGATCAGATAGACATCTCCGAGTACATG 2017
 Qy 640 GluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyroValGlyHisGluGlu 659
 Db 2018 GAACAACAATCTGTTTCCCGCTGATTGCTGCTCCACCGGTTATTTGGGCGATGATGAG 2077
 Qy 660 GlyGlyGlnLeuThrGluAlaValArgArgArgProTyroCysValIleLeuPheAspGlu 679
 Db 2078 GGAGGACAACTTACTGAAGCTGTGTAGGAGGGGGCTTACAGTGTGTGCTATTGATGAA 2137
 Qy 680 ValGluLysAlaHisAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArg 699
 Db 2138 GTTGAGAAAGCCCATCTCGAGTGTGTTAATACATGCTTCAAGTCTTGATGATGAGGAG 2197

700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSer 719
 2198 TTAACAGATGGTCAAGCGCCGACAGTTGATTTACCAACTCGTGATTTATGACTTCA 2257
 720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLeuValMetGluValAla 739
 2258 AACTTGGGAGCAGATATCTGTTGCTGATTAATGGGCAATGTACCATGGAGCAGCT 2317
 740 ArgAspCysValMetArgGluValAlaGlyHisPheArgProGluLeuLeuAsnArgLeu 759
 2318 CCGTGAATCGTATCGACGAGGTGGAAAGCAGTTTAAGCCCGACTCTGTAATCGGTG 2377
 760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
 2378 GATGAGATTGTTGTTGATCTCTGTCACAGCAGTGTGAGCAAGTATCGCGCTAC 2437
 780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
 2438 CAGATGAGGAGCTTGCACACTACGGCTGGCTGAGAGGGGTATTGCATTGGCGCTTACTGAG 2497
 800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyValaArgProIle 819
 2498 CGAGCTCCAGATGTATATCTACAGAGATTATGACCCGGTTTATGGTGCAAGACTATT 2557
 820 ArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
 2558 AGGAGATGTTGGAGAGGAAGTGGTGACCGAGTATCCAGATGCTGTGAAGGAGGAG 2617
 840 IleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuValTyr 857
 2618 ATTGATGAGAATCTCAACGGTTTACATAGATGCTGGGGTCAAGCGGAAAGATCTACCTAC 2677
 858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeu 876
 2678 AGGTGGAGAAAGTGGAGGGCTTGTGAATGCTGCACCGGCAAAATCTGATATATG 2737
 877 IleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMetArg 896
 2738 ATTCACTTCTTAATGGTCCC---AGGAGTGATGCTGTCCAGCAGTCAAGAGATGAGG 2794
 897 IleGluGluIleGluAspAspAspGluGlu 907
 2795 ATTGAAGAAATTGAAGATGACGAAATGGAAGAT 2827

MULT 7

566070

AAI66070 standard; DNA; 3084 BP.

AAI66070;

14-JAN-2002 (first entry)

Zea mays heat shock protein HSP101 gene (GenBank: AF133840).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana; ds.

Zea mays.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.
 DR P-PSDB; AAM51669.
 XX Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -
 PS Claim 4; Page -: 91pp; English.
 XX The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 CC Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.

XX SQ Sequence 3084 BP; 670 A; 874 C; 1057 G; 483 T; 0 other;

Alignment Scores:

Pred. No.:	1.45e-278	Length:	3084
Score:	3883.50	Matches:	772
Percent Similarity:	92.88%	Conservative:	76
Best Local Similarity:	84.56%	Mismatches:	58
Query Match:	85.35%	Indels:	7
DB:	22	Gaps:	6

US-09-812-350-17 (1-911) x AAI66070 (1-3084)

Qy 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 145 ATGAATCCGACAACTTCCACCAAGACGAGCGCATCGTGGGGCGCACGAGATT 204
 Qy 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 Db 205 GCGGTGGAGGCCGCCACGCGCAGCTCACGCCCTGCACCTGCCGCCGCTGGTGGCTGC 264
 Qy 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
 Db 265 GACAGGGCGGCATCTCGCGGACAGGCATTCGCGGGGGCGTCCGGGGGCGACGAGCGGCC 324
 Qy 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 Db 325 GGGGACTCGTTCGAGCGCGTGTGAACAACTCGCTCAAGAAGCTGCGCTCGCAGTCCCG 384
 Qy 79 ProProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 Db 385 CCGCGGACTCCCTCCCGGCTCCACGGCACTCATCAAGTCAATCCCGGGCGGAGTCC 444
 Qy 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 Db 445 GCGCAGAAAAACGCGGGGACTCTCACCTCGCGCTGACCAAGCTGCTGCTGCTGCTGCTC 504
 Qy 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValIglyValAlaThrAlaArgValLys 138
 Db 505 GAGGACTCGCAGATCTCCGACTCCCTCAAGGAGCGCGGTGTCCCGCGCGGGTGGCG 564
 Qy 139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 Db 565 GCCGAGCTTGAGAACTCCCGCGGGGAGGGCGCGCTGGAGTCCCGCTCGGGGAGAC 624
 Qy 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
 Db 625 ACCAACTTCCAGGGCTCAAGACATACGCGCGGGAACCTGTTGAGCAGCGCGGAGCTG 684

179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
685 GACCCCGCTCATCGCCGCGACGAGAGATCCGCGCTGCTGCGCATCTCTCGCGCCG 744
199 ThrLysAsnAspProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
745 ACTAGAACACCCCGCTCTCATCGCGAGCCGCGGCTTGGCAAGACGCGCTGCTGGAG 804
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
805 GGCCTCGCGACGCGATCTTCGCGCGAGCGTCCAGTAACCTCTCGAGCTCGCGCTC 864
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGluPheGluGlu 258
865 ATCGCGCTCGACATGCGCGCTGCTCGCGCGCGCAAGTACCGCGCGAGTTCGAGGAG 924
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
925 CGGCTCAAGCGCTGCTCAAGAGGTGGAGAGCGCGGAGGGAAGGTCAATCTCTTCATC 984
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
985 GACGAGATACCTCGTCTCGCGCGCGGCGAGCGAGGGTTCATGGACGCGCGCCAC 1044
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
1045 CTGTTCAAGCCCAATGCTGGCGAGGCGACAGCTCAGGTGCATCGCGCGCACACGCTGGAG 1104
319 GluTyArgLysTyValGluLysAspAlaAlaPheGluArgArgPheGlnValTy 338
1105 GAGTACCGCAAGTACGTGGAGAGAGCGACGAGCTTCGAGCGCGGTTCCAGCAGGTGTT 1164
339 ValAlaGluProSerValProAspThrIleLeuArgGlyLeuLysGluLysTy 358
1165 GTCGCGAGCGCGAGCGTCCGCGACACCGTCACTTCAGCGGACTCAAGGAGAGTAC 1224
359 GluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
1225 GAGGCGCACCATGGCGCTGAGGATCCAGACCGCGCGCTCGTGGCGCGACAGCTATCC 1284
379 AlaArgTyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1285 GCGAGGTACATCATGGTTCGCGACCTGCTGCAAGCCATAGACTTGTGGACGAGGCC 1344
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuArg 418
1345 TCGCGCAATGTGAGGTGCGAGCTCCACAGCGCGAGCGAGGAGATTGATACCTGGAGAGG 1404
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1405 AAGAAATCCAGCTTGAGGTTGAGTCCACCGCTCGAGAGGAGGAGGAGCAAGCGCCAGC 1464
439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458
1465 AAAGCCCGCTGATCGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1524
459 LeuThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGln 478
1525 CTGACCATCAAGTACAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1584
479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyArgTyArgLeuAlaArg 498
1585 CCGCGCGAGGAGCTCAAGTTCACTCCCTCGAGGAGGCGCGCGCGCGCGCGCGCGCGCT 1644
499 AlaAlaAspLeuArgTyGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
1645 GTGGCGGACCTCAAGTACGCGCGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1704
519 GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAla 538
1705 ---AGCGAAACAGGGGAGAACCTGATGCTACCGGAAACCGTCCGCGCGCTCAAGAAATGCA 1761
539 GluValValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGlu 558

1762 GAGGTGCTGAGCGCTTGGACGGGTATTCCAGTGATCCCGCTTGGCCAGAACGACAGGAG 1821
559 ArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaVal 578
1822 AGGTGCTGCGCTTGGCTTGGCTGACAGGCTTCCACAGAGGGTGGTTCGCGCCACAGAGGCTGTG 1881
579 AsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnPro 598
1882 AGCCCGCTCGCAGAGGGCGGTGCTGAGGTCCAGGGCGCGCTTTCGCGAGGCCACACAGGCC 1941
599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAla 618
1942 ACTGCTCGTCTCTGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2001
619 LeuAlaGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTy 638
2002 CTAGCCGAAACAGCTGTTTCGACGACGAGAACCTTCTTGTCCGATCGACATGTCGGAGTAC 2061
639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyValGlyHisGlu 658
2062 ATGAGCAGCAGCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2121
659 GluGlyGlyGlnLeuThrGluAlaValArgArgArgProTyCysValIleLeuPheAsp 678
2122 GAGGTGCGCGAGCTGACTGAACAAGTGAGGAGGAGGCGGTACAGCGGTGATCTCTGTTCCGAC 2181
679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly 698
2182 GAGTTCGAGAGGCCCATGTCGCGCTGTTCAACACCTGCTCCAGGTCCTTCGACGACGCG 2241
699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThr 718
2242 AGTTGACGATGCGCAAGCGCAGCGCGTGGACTTCAGGAACACCGGTGATCATCATGACA 2301
719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
2302 TCGAACCTCGCGCGCGCAGCAGCTCTCTGCTGGATGTTGGCGAAGAACTCCATGAAGGTC 2361
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2362 GCTCGCATCTGCTCATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2421
759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2422 CTCGACGAGATCTGATCTGATCTCTGCTGCTCCACGACGAGCTGAGGAGGAGGAGGAGGAG 2481
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2482 CTTGATGAAGGATGTCGCGCTTTCGCGAAGGCGCATCTCTGCTGCTGCTGCTGCTGCTGCTG 2541
799 AspAlaAlaLeuAspTyIleLeuAlaGluSerTyArgProValTyArgValAlaArgPro 818
2542 GACGCGCATTTGACATCTCTGCTCTCTTTCGATCCGCTGATCCGCGCGCGCGCGCGCG 2601
819 IleArgTrpMetGluLysValValThrGluLeuSerLysMetValValArgGlu 838
2602 ATCAGGAGGTGGATCCGAGAGAGGGTGGTGGAGCGAGCTCTCGAGATGCTGATCCAGGAG 2661
839 GluIleAspGluAsnSerThrValTyIleAspAlaGlyAlaGly-----AspLeuVal 856
2662 GAGATCGACGAGAACTGCACGGTCTATCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2721
857 TyArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
2722 TACAGGTTGACCGGAGCGCGCTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2781
876 LeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMet 895
2782 CTGATCCAGTCCCGCAACAGCTCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2841
896 ArgIle---GluIleGluAspAspAspAsnGluGlu 907

[illegible]

399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
1294 TGCGCAATGTGAGGGTCACTTGCAGCCAGCGCTGAGAGATCGAACCTCGAGAG 1353
419 LysArgMetGlnLeuGluIleGluHisAlaLeuGluArgGluLysAspLysAlaSer 438
1354 AAGAGATCCAGTGTGAAGTCCAGCTCCATCCCTCGAGAGAGAGAGCAAGAGCTAGC 1413
439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458
1414 AAAGCTCGCTAGTTGATGTGAGAGAGAGTGTGACCATCTGAGAGACAAGCTCGACCG 1473
459 LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGln 478
1474 CTGCAGATGAAGTATCGCAAGAGAGAGAGGATCGACGAGATCAGAGCGCTGAACGAG 1533
479 LysArgGluGlnLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArg 498
1534 CGCGCGAAGAGCTGCAGTTCCTCTGAGAGAGCGCGAGCGCGGATGATTTGGCCCGC 1593
499 AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
1594 GTGCTGATCTCAGATACGGTGTCTGAGAGAGTTCACGCTGCCATTTGCTAAGCTGGAG 1653
519 GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAla 538
1654 GGTGAGACCGGC--GAGAAATTGATGTAAACAGAGACTGTGCGCCCGGACCATTTGCT 1710
539 GluValValSerArgTyrThrGlyIleProValThrArgLeuGluGlnAsnGluLysGlu 558
1711 GAGTGTGTGAGCCCTGGACTGTGATTCCTGTCACCGAGGTTCGACAGATGAGAGGCG 1770
559 ArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaVal 578
1771 AGGCTGATCGGGCTGGCAGATCGACTCATCAGAGGCTGTGTGACAGATGAGGCGGTC 1830
579 AsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnPro 598
1831 AATCAGTTGGAGAGGCTGTTCTACGCTCGAGGCTGCGCTTGGACCGCCGACGAGCT 1890
599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAla 618
1891 ACTGTTTCATTCCTTCCTTGGACCGACCGGCTGTCGGAACCAACCGAGCTCGCAAGGCT 1950
619 LeuAlaGluGlnLeuPheAspGluAsnLeuLeuValArgIleAspMetSerGluTyr 638
1951 CTAGCTGAGCAGCTGTTGATGACGAGAACTGCTGCTCGCATCGACATGCTGTAATAC 2010
639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGlu 658
2011 ATGAGCAGCATTCGGTTGCGCGCTAATCGAGCCCACTCGATATGTTGTCATGAA 2070
659 GluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAsp 678
2071 GAAGGGGGGAGCTGACCGAGCAAGTGAAGAGGAGGCAATACAGTGTTCATCTTCGAC 2130
679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly 698
2131 GAGTTGAGAGGCGCATGTGGCGGTGTTAACTCTGCTCCAGTCTTGGAGCATGGG 2190
699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThr 718
2191 CGTTTGACCGAGCGGCAAGGAGGAGCGGTTGATTTAGGAACACGGTGTATCATCATGACC 2250
719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
2251 TCAACCTTGGCGGAGACCTCTCGCGGAAATGTTGGTGGCAAT---TCGATGAAGTT 2307
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2308 GCTCGTGTATCTGTCATGAGAGGTGAGAGGCAATTTCCGCGGAGCTGCTGAACCGT 2367

759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2368 CTGAGCAGATGTCTATCTTCCACCTCTGTGATGAGCAGCTCGGAGGTCGCTCGG 2427
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2428 CTTTCAGATGAAGATGTGCGAGTCCGCTCTTCCGAGAGGCGGCTTCTCTGCGCGTCACC 2487
799 AspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818
2488 GAGCGCGCTCGACGTCATCTCTGCTCTTACGATCCGCTCTATGCGCCGACGCGCA 2547
819 IleArgTyrMetGlnLysLysValValThrGluLeuSerLysMetValValArgGlu 838
2548 ATCCGAGATGATCGAGAGAGGATGATGACGAGCTCTCCAGATGTTGATCCGCGAG 2607
839 GluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuVal 856
2608 GAGATCGACGAGAACTCCACGCTGTACATCGAGCTGCGCCCGACCAAGCAGCTGACC 2667
857 TyrArgValGluSer---GlyGlyLeuValAlaAspAlaSerThrGlyLysLysSerAspVal 875
2668 TATGCGCTCGACCAAGCAGCGGCTGTGAACGCGCGCACGCGCCACCAAGTCCGACATC 2727
876 LeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMet 895
2728 CTGATCCAGTTCCTAGCGAGCTGTTGGGCGGATGCGCGCACGCGCTGAGAGATG 2787
896 ArgIle-----GluGluIleGluAspAspAsnGluGlu 907
2788 AAGATCATGACGAGCAGCGAGCGAGGTGGACGACGATCGAGGAA 2829
RESULT 9
AA166075
ID AA166075 standard; DNA; 2821 BP.
AC AA166075;
DT 14-JAN-2002 (first entry)
XX
DE Triticum aestivum 101kDa heat shock protein gene (GenBank: AF083344).
KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
KW Arabidopsis thaliana; ds.
XX
OS Triticum aestivum.
XX
PN WO200170929-A2.
XX
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-US08836.
XX
PR 20-MAR-2000; 2000US-190769P.
PR 18-APR-2000; 2000US-198116P.
XX
PA (ARCH-) ARCH DEV CORP.
XX
XX Lindquist S, Queitsch C, Vierling B;
XX WPI; 2001-639123/73.
DR P-PSDB; AAM51661.
XX
PT Transgenic plants with improved heat stress tolerance, useful for
PT producing animal feed, oil and synthetic products -
XX
PS Claim 4; Page -; 91pp; English.
XX
CC The invention relates to a transgenic plant, comprising a genetic
CC construct comprising a promoter operatively linked to a nucleic acid
CC sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has

increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 2821 BP; 524 A; 902 C; 1035 G; 360 T; 0 other;

Alignment Scores:

Seq. No.: 1.12e-271 Length: 2821
 Score: 3790.00 Matches: 752
 Percent Similarity: 91.09% Conservative: 86
 Mismatches: 81.74%
 Indels: 12
 Gaps: 7

us-09-812-350-17 (1-911) x AAT66075 (1-2821)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 65 ATGAAACCGGCAACCTTCAGCGCAACAGGAGCGCGTGGTGGCGCGCAGCGGCG 124
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40
 125 GCGTCGGAGCGCGCGCACCGCGCAGATCAACCGCTTGCACCTGGCGCGCGCGTGGCGGCG 184
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla 59
 185 GACAAGTCGGGCGATCTCGCGGAGCGCGTGGCGGCGCGTGGCGGCGGATGCGTCCGCG 244
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 245 GGGGACTCGTTCAGGCGCGTGTCTCGCGCGCGCGCTCAGGAAGCTGCGTCCGAGTCCGCG 304
 79 ProProAspAlaIleProAlaSerSerSerSerSerSerSerSerSerSerSerSerSer 98
 305 CCGCGCGACTCGCGCGCGCTTCAAGCGCGCTCATCAAGCCATCCCGCGCGCGCGTCCG 364
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 365 GCGCAGAAAGCGCGGAGCTCGACCTCGCGCTCGACCGAGTGTCTATGGGCTCTCTC 424
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 425 GAGGACCGCGAGATCGCGACTCGCTCAAGAGGCGCGCGTGTCCGCGCTCGCGGCGCGC 484
 139 SerGluValGluLysLeuArgGlyLysGlu---GlyLysLysValGluSerAlaSerGly 157
 485 GCGGAGCTCGACAGCTCCGCGCGGCGCAACTCGCGCAAGTTCGAGTCCGCTTCGCG 544
 158 AspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLys 177
 545 GACACCACTTCAGGCGCTCAAGAGCTACGCGCGCGACCTCTCGTTCGAGGTGGCGCGCAAG 604
 178 LeuAspProValIleGlyArgAspGluLeuIleArgArgValValAlaArgIleLeuSerArg 197
 605 CTCGACCCCGTATCGGCGCGCGAGAGATCCGCGCGCTGTGCGCATCTCTCGCGC 664
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 665 CGCACCAAGAACACCCCGCTCATCGCGGAGCGCGCGTTCGGCAAGCGCGCTCGTCT 724
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 725 GAGGGCTTCGCGCGCGCGCGTGTGCGCGCGCGTGTGCGCGCGCGTGTGCGCGCGCG 784
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGlyGluPheGlu 257

Db 785 CTCGTCGCGCTCGACATGGCGCGCTCGTCCGCGCGCAAGTACCGCGCGAGTTCCGAG 844
 Qy 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 Db 845 GAGCGCTCAAGCGCGTGTCTCAGGAGTGGAGAGCGCGCAGCGCAGAGTTCATCTTTC 904
 Qy 278 IleAspGluLeuHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 Db 905 ATCGACGAGATACACCTCGTCTCGCGCGCGCGGAGCGGAGGCTCATGGAGCGCGCC 964
 Qy 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317
 Db 965 AACCTCTTCAAGCCCATCTCGCGCGCGCGCGTCTCGGTGTCATCGCGCGCACACGCTG 1024
 Qy 318 GluGluTyrglyLysTyrglyValGluLysAspAlaPheGluArgArgPheGlnGlnVal 337
 Db 1025 GAGGAGTACCGCAAGTACGTTCGAGAGGAGCGCGCTTCGAGCGAGCGTTCGAGCAGGTG 1084
 Qy 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
 Db 1085 TACGTGCGCGAGCGCGTCTCGCGCACCATCAGCATCTCTCGAGGGCTCAAGGAGAG 1144
 Qy 358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 Db 1145 TACGAGGCGCACCGCGCGTGTGCGATCCAGGACCGCGCATCTGTCGCGCGCGCGCTC 1204
 Qy 378 SerAlaArgTyrglyThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 Db 1205 TCGGCGAGGTACATCATGGCGCGCACCTTCGCGGACCAAGCCATCGACCTGGTGGAG 1264
 Qy 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 Db 1265 GCGTCGCGCAACGTCAGGTCGAGTTCGAGACAGCGCGCGGAGGATCGCAATCTGGAG 1324
 Qy 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAspLysAla 437
 Db 1325 CGAAGCGGATTCAGCTGGAGTGGAGTGCAGCGCTGAGAGGAGGAGGAGGAGGAGGAG 1384
 Qy 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 Db 1385 AGCAAGCGCGCGTGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 1444
 Qy 458 ProLeuThrMetLysTyrglyArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 Db 1445 CCGCTGACCATGAAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1504
 Qy 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrglyAspLeuAla 497
 Db 1505 CAGCGCGCGGAGGAGTTCAGTTCAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1564
 Qy 498 ArgAlaAlaAspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 Db 1565 CCGCTCGCGGACCTCAAGTACGCGCGCTTCAGGAGATCGACCGCGCATTCGCAAGCTG 1624
 Qy 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 Db 1625 GAGCGCGAGACCGCGC---GAGAACCTGATCTCAGGAGACCGTCCGCGCGCGCGAGATC 1681
 Qy 538 AlaGluValValSerArgTyrglyIleProValThrArgLeuGlyGlnAsnGluLys 557
 Db 1682 CCGGAGTGTGTGAGCGGTGAGCGCGCATCCCGGTGACCGCGTTCGCGGAGGAGGAGGAG 1741
 Qy 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 Db 1742 GAGCGCTGTGTGCGGATCGCGGACAGGCTGCACAGGAGGAGTGTGTGTGTGTGTGTGTGTGT 1801
 Qy 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 Db 1802 GTGAACCGCTCGCGGAGGCGGTGTGAGTTCAGGCGCGGCGTGTGTGTGTGTGTGTGTGTGT 1861
 Qy 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617

1962 CCCACCGGGTCTTCCTCTCTGGCCGACCGCGTGGCCAGACAGAGCTCGCCAG 1921
618 AlalaGluGlnLeuPheAspGluAsnLeuValArgLeuMetSerGlu 637
1922 GCGCTCGCCGAGACCTGTCGACGACGAGACCTGCTCGCATCGACATGTCGAG 1981
638 TyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProGlyTyrValGlyHis 657
1982 TACATGAGCAGCAGCTCGCTCGCTCATCGCGCCGACCCAGCGCTATGTCGACAC 2041
658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
2042 GAGAGGCGCGGACGTGACGAGCAGCTGAGGAGGAGCGGTACAGCGTATGCTGTC 2101
678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 697
2102 GACGAGTGGAGAGCGCAGCTGCGGTGTTCACACGCTCTCCAGGTGCTGGACGAC 2161
698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMet 717
2162 GCGCGGTGACCGACGCGGACGCGGACGCTGACCTCCGGAACACCGTGATCATG 2221
718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
2222 ACCTCCAACTCGCGCGGACGACCTCTCGCGGATGTGGCAAGAACTCCATGAAG 2281
738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
2282 GTGCGCGCGACCTGTCATGACGAGGTGAGGAGGACCTTCAGCGCGAGCTGCTGAAC 2341
758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
2342 CGGTGACGAGAGTGTGATCTTCACCGCTGTCACGAGCAGCTGAGAAAGTGGCC 2401
778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuVal 797
2402 CGCTGACAGTAGAGGACGTGGCGCTCGCGTCCGAGCGCGGCTGCGACGTGGCGTC 2461
798 ThrAspAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArg 817
2462 ACCGACCGCGCTGGAGCTCATCTGTCGTGGCTTACGATCCAGTGTATGGGCTCGG 2521
818 ProLeuArgArgTrpMetGluLysValValThrGluLeuSerLysMetValValArg 837
2522 CCGATCAGGAGTGGATCGAGAGAGGCTGTGACGAGCTGTCCAGATGCTGATCCAG 2581
838 GluGluLeuAspGluAsnSerThrValTyrLeuAspAlaGlyAlaGlyAsp---LeuVal 856
2582 GAGGAGATCGACGAGATTCACATGTCATCGCCGCGGACGAGGACGAGCTGGCC 2641
857 TyrArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
2642 TACCGGTGGACCGCAGCGCGGCTGTGTGAACCGCCGAGAGCGGCGAGGTCCGACATC 2701
876 LeuIleHisIleAlaAsnGlyPro-----LysArgSerAspAlaAlaGlnAla 891
2702 CTCATCAGGTGCGCAACCGCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2761
892 ValLysLysMetArgIleGluGluLeuGluAspAspAsnGluGluMetIleGluAsp 911
2762 GTGAAGAAGATGAGGGTC-----ATGAGAGCGGGGACGAGGCGGCGGACGAGAT 2815

MULT 10

66073

AAI66073 standard; DNA; 3942 BP.

AAI66073;

14-JAN-2002 (first entry)

Phaseolus lunatus heat shock protein HSP100 gene (GenBank: AF203700).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
KW Arabidopsis thaliana; ds.
OS Phaseolus lunatus.
XX WO200170929-A2.
PD 27-SEP-2001.
XX 20-MAR-2001; 2001WO-US08836.
XX 20-MAR-2000; 2000US-190769P.
PR 18-APR-2000; 2000US-198116P.
XX (ARCH-) ARCH DEV CORP.
XX Lindquist S, Queitsch C, Vierling B;
XX WPI; 2001-639123/73.
DR P-PSDB; AAM51671.
XX Transgenic plants with improved heat stress tolerance, useful for
XX producing animal feed, oil and synthetic products -
PS Claim 4; Page -; 91pp; English.
XX The invention relates to a transgenic plant, comprising a genetic
XX construct comprising a promoter operatively linked to a nucleic acid
XX sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
XX family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
XX increased stress tolerance, especially to heat. The plant is a cereal,
XX grass, ornamental plant, crop plant, food plant, oil-producing plant, a
XX synthetic product-producing plant, an environmental waste absorbing
XX plant, an alcohol plant, a medicinal plant, a recreational plant and/or
XX an animal feed plant. In particular, the transgenic plant is cotton,
XX canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
XX Arabidopsis thaliana. The plants may be used to produce animal feed,
XX alcohol, crop, oil, medicine or a synthetic product.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained from GenBank using the Accession Number
XX reference provided in the specification.
SQ Sequence 3942 BP; 834 A; 1111 C; 1236 G; 760 T; 1 other;
Alignment Scores:
Pred. No.: 1-248-261 Length: 3942
Score: 3657.00 Matches: 770
Percent Similarity: 78.2% Conservative: 76
Best Local Similarity: 71.2% Mismatches: 60
Query Match: 80.3% Indels: 177
DB: 22 Gaps: 11
US-09-812-350-17 (1-911) x AAI66073 (1-3942)
QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 436 ATGAATCCGCGACAACTTCCACCAAGACGAACGAGGCGCATCGTGGGGCGCGACGAGATT 495
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
DB 496 GCGGTGGAGCGCGCGCACGCGCAGCTCACCGCGCTGACCTGGCCGCGAGTCTGCTGCTCG 555
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
DB 556 GACAAAGGCGGCATCTCGCGCGAGCCATCACGGGGGCGTGGGGGGCGACGAGCGGCC 615
QY 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
DB 616 GGGGACTCGTTCGAGCGGTGCTGAACCAACTCGCTCAAGAAGCTGCGCTCGAGTCCCG 675
QY 79 ProProAspAspIleProAlaSerSerSerLeuIleValIleArgAlaGlnAla 98
DB 676 CCGCGGAGTCTCGGCTTCCGCGGTACCGCGCTGATCAAGGTATCCCGCGGCGCGAGTCC 735

99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuLeuMetGlyLeuLeu 118
 736 GCGCAGAGAAACGCGGGAGACTCGACCTCGCGTGCACCGAGTGTGTCGCGCTGCTC 795
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 796 GAGGACTCGCAGATCTCCGACTCCGACTCAAGAGAGCGCGGTGTCCGCGCGCGGTGCGC 855
 139 SerGluValGlnLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 856 GCGGAGCTTGAGAGCTCCGCGCGGGAGCGCGCGGTGAGGTCCGCGTGGGGAT 915
 159 ThrAsnPheGlnAlaLeuLysThrTrpGlyArgAspLeuValGlnAlaGlyLysLeu 178
 916 ACCAATCTTCAGCGCTCAAGACATACGCGCGGACCTCGTCGAGCAGCGCGGAGCTT 975
 179 AspProValIleGlyArgAspGluGlnIleArgArgValValArgIleLeuSerArgArg 198
 976 GACCCCGCTCATCGCGCGCGAGAGAGATCCGCGCGTGTGCGCATTCCTCGCGCGCGC 1035
 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
 1036 ACCAAGAAATACCCCTCTCATCGCGAGCGCGCGGTGGCAAGACGCGCGTGTGGAG 1095
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 1096 GCGCTCGCGCAGCGCATCTCGCGCGAGCGTCCGCGAGTAACTCTCGACCTCGCGCTC 1155
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTrpArgGlyGluPheGluGlu 258
 1156 ATCGCGCTCGACATGGCGCGCTCGTGGCGCGCGCGCAAGTACCGCGCGAGTTCGAGGAG 1215
 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
 1216 CGGCTCAAGCGCTGTCAAGAGAGTGGAGAGCGCGAGCGAGTCAATCTCTTCATC 1275
 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
 1276 GACGAGATACACTCTCTCGCGCGCGCGCGAGCGAGCGGTTCATGAGCGCGCAAC 1335
 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 1336 CTGTTCAAGCAATGCTGGCGAGGACAGCTCAGTGTGATCGCGCGCCACCGCTGGAG 1395
 319 GluTrpArgLysTrpValGluLysAspAlaAlaPheGluArgArgPheGlnValTrp 338
 1396 GAGTACCGCAAGTACGTGGAGAGGACGCGCTTCGAGCGCGGTTCAGCAGGTGTTTC 1455
 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyTrp 358
 1456 GTCCGCGAGCGGAGCGTCCGACACCGTCAAGCATCTCTGAGGCGCTCAAGAGAGTAC 1515
 359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 1516 GAGGGACCATATGGCGTGGAGATCAGAGCGCGCGCTCTGTCGCGGACAGCTCTCC 1575
 379 AlaArgTrpIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
 1576 GCGAGGTATCATATGGTTCGACCTCGCTGACAAAGCCATAGACCTGTGTGACGAGGCC 1635
 399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
 1636 TGCAGCAATGTAGGGTGCAGCTCGACAGCCAGCGAGGAGATTCATTAACCTGGAGAGG 1695
 419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAspLysAlaSer 438
 1696 AAGAGATCCAGCTTGAGGTGAGCTCCAGCGCTCGAGAGAGAGAGAGAGAGAGAGAG 1755
 439 LysAlaArgLeuIleGluVal----- 445
 1756 AAAGCCCGGCTGATTGAGGT-GGGTGTCTCGAGCACTGAATTTCTCTCAACAAAGTCCTC 1814

445 ----- 445
 1815 TTGTCGTGATGTCCTGTTGTTGAACTTTTGATCTAACTGTCGCTTAACTGTGGCCATTGTT 1874
 446 -----ArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTrp 463
 1875 CAGGTCAGGAGAGGAATTTGGACGATCTGAGGACAAAGTGCAGCCCTGACCATGAGTAC 1934
 464 ArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeu 483
 1935 AGGAGAGAGAGAGAGATTTGATGATCAGGAGCTGAAGCAGAGCCCGCGAGGAGCTC 1994
 484 MetPheSerLeuGlnGluAlaGluArgArgTrpAspLeuAlaArgAlaAlaAspLeuArg 503
 1995 CAGTTTCACTTCAGGAGGCGGAGCGCGGATGACCTGCGCGTGTGGCCGACCTCAAG 2054
 504 TyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGlu 523
 2055 TACGCGCGCTTCCAGGAATTCGACGCTGCTATCTCAAGCTGGAG--AGCGAAACACAGG 2111
 524 GluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu----- 539
 2112 GAGAACCTGATGCTCACCGAAACCGTGGCGCTGACAAATTCGAGAGGTATGTTATTAT 2171
 539 ----- 539
 2172 TCTTTGTTACCGTCACAAAATTTTGCAGAGCAAGTGCAGATTTTGGCGATCTCTCCTA 2231
 540 -----ValValSerAr 543
 2232 GTAGAGTAGTCGTCACGCTGCTGAAATGTTGTTCTATATGGGCGAGTGTGTGAGCG 2291
 543 qTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLe 563
 2292 TTGACGCGGTATTCAGTGACCCGCTGGCCAGACGACAGAGAGGCTGTGTGGCT 2351
 563 uAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerG 583
 2352 GGTTCAGAGGCTTCACAGAGGCTGTGGCGACAGAGAGCTGTGAGCGCGCTCGCAGA 2411
 583 uAlaIleuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLe 603
 2412 GCGCGTCTGAGGTCGAGGCGCGCTTGTGGCAGGCCCAACAGCCCACTGGCTGCTTCT 2471
 603 uPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLe 623
 2472 CTCTCTGGTCCGACTGGCGTGGGAAACTGAGCTGGCCAGGCCCTAGCCGACAGCT 2531
 623 uPheAspAspGluAsnLeuValArgIleAspMetSerGluTrpMetGluGlnHisSe 643
 2532 GTTCGACGACGAGAACCTTCTTGTCCGATCGATGTCGAGTACATGAGGAGCAGCTC 2591
 643 rValSerArgLeuIleGlyAlaProGly----- 653
 2592 GTTGGCGCGCTCATCGAGCACCACTCTGG- GTAAGTAGCAGAAAATGATGATCTTCG 2650
 653 ----- 653
 2651 TATTTTAACTGCGAATTCGATGAACTCTTGTCTGACAGTGGCGCGGACCTGTGACAG 2710
 654 -TyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgArgProTrpC 673
 2711 CTAGTCCGCGCATGAAGAGGTCGCGAGGCTGACTGAACAGTGAAGAGGAGGCGGTACAG 2770
 673 sValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuG 693
 2771 CGTGATCTGTTTCGACGAGGTTCGAGAGGCGCATGTGCGCGTGTCAAACCTCTGCTCA 2830
 693 nValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgHisSe 713
 2831 GGTCTCTCGACGACGCGAGGCTGACGAGTGGCAGGAGGAGGAGGAGGAGGAGGAGGAG 2890
 713 rValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLy 733

[illegible]

DATA66074.

14-JAN-2002 (first entry)

Triticum aestivum heat shock protein 101 gene (GenBank: AF097363).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato; *Arabidopsis thaliana*; ds.

Triticum aestivum.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08936.

PR	20-MAR-2000; 2000US-190769P.
PR	18-APR-2000; 2000US-198116P.
XX	
XX	(ARCH-) ARCH DEV CORP.
PA	
XX	
PI	Lindquist S, Queitsch C, Vierling E;
XX	
XX	WPI; 2001-639123/73.
DR	P-PSDB; AAM51660.
DR	
XX	
XX	Transgenic plants with improved heat stress tolerance, useful for
PT	producing animal feed, oil and synthetic products -
XX	
XX	Claim 4; Page -; 91pp; English.
PS	
XX	
CC	The invention relates to a transgenic plant, comprising a genetic
CC	construct comprising a promoter operatively linked to a nucleic acid
CC	sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
CC	family amino acid sequence (AAM51651-AA51671). The transgenic plant has
CC	increased stress tolerance, especially to heat. The plant is a cereal,
CC	grass, ornamental plant, crop plant, food plant, oil-producing plant, a
CC	synthetic product-producing plant, an environmental waste absorbing
CC	plant, an alcohol plant, a medicinal plant, a recreational plant and/or
CC	an animal feed plant. In particular, the transgenic plant is cotton,
CC	canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
CC	Arabidopsis thaliana. The plants may be used to produce animal feed,
CC	alcohol, crop, oil, medicine or a synthetic product.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained from GenBank using the Accession Number
CC	reference provided in the specification.

Sequence 3942 BP; 834 A; 1111 C; 1236 G; 760 T; 1 other;

Alignment Scores:	
Pred. No.:	1.24e-261
Score:	3657.00
Percent Similarity:	78.56%
Best Local Similarity:	71.23%
Query Match:	80.37%
DB:	22
	Gaps: 11
	Indels: 177
	Mismatches: 60
	Conservative: 76
	Matches: 770
	Length: 3942

US-09-812-350-17 (1-911) x AAI66074 (1-3942)

QY	1	MetAsnProGluLysPheThrHisLeuThraAsnGluThrIleAlaThrAlaHisGluLeu	20
Db	436	ATGAATCCGGCAACACTTCACCAACAAGACGAGCGCATCTGGGGCGCAGAGATT	495
QY	21	AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer	40
Db	496	GGCGTGGAGGCGCGCACGCGACACTACGCGCGTGCACTGGCCCGCAGTGTGGCTGGCG	555
QY	41	AspProThrGlyIlePheProGlnAlaIleSerSerAla--GlyGlyGluAsnAlaIle	59
Db	556	GACAGGGCGGCATCTCTGCGCAGCGCATACGGGGGCGTCTGGGGGGCGACGAGAGCGGCC	615
QY	60	---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro	78
Db	616	GGGGACTCGTTCGAGCGCGTCTCAACAACACTCGCTCAAGAAGCTGCGCTGCAGTCCCCG	675
QY	79	ProProAspAspIleProAlaSerSerSerIleLysValIleArgArgAlaGlnAla	98
Db	676	CGCGCGGACTCGTTTCGGCGCTCTACGCGCGTCTCAAGGTCTATCCGCGCGCGCGCAGTCC	735
QY	99	AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu	118
Db	736	GGCGAGAAGAAACGGGGGACTCGCACTCGCGTGCACAGTGTCTGTCTGGCCCTGCTC	795
QY	119	GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys	138
Db	796	GAGGACTCGCAGATCTCCGACTGCTCAAGGAGCGCGCGTGTCTCCGCGCGCGGTGGCG	855
QY	139	SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaArgGlyAsp	158

856 GCCAGAGCTTCAGAAAGCTCCGCGCGGGAGGGCGCGCGGTGAGTCCGCGTGGGGGAT 915
159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
916 ACCAATTCAGCGGCTCAAGATACGCGCGGAGCTCTGTCAGCAGCGCGGAGCTT 975
179 AspProValIleGlyArgAspGluIleArgArgValValArgIleLeuSerArgArg 198
976 GACCCGCTCATCGCGCGCAGCAGAGAGATCCGCGCGTCTGTCGCATTCCTCGCGCGC 1035
199 ThrLysAsnProValIleGlyGluProGlyValGlyLysThrAlaValGlu 218
1036 ACCAAGATTAACCCCGCTCATCGCGAGCGCGCGTGGCAAGACGCGCTCGTGGAG 1095
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
1096 GGCCTCGCGCAGCGCATCTTCGCGCGAGCTGCCAGTAACCTCTCGAGTCCGCTC 1155
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
1156 ATCGCGCTCGACATGGCGCTCTCAGGAGGTGGAGAGCGCGCAAGTACCAGCGCGAGTTCGAGGAG 1215
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
1216 CGCTCAAGCCCGTCTCAGGAGGTGGAGAGCGCGCGAGGTCAATCTCTTCATC 1275
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
1276 GACGAGATACACCTCTCTGCGCGCGCGCAGCAGCGAGGGTTCATGACGCGCGCAAC 1335
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
1336 CTGTTCAAGCCCATGCTGCGCAGGAGGACAGCTCAGGTGCATCGCGCGCACCACTCGTGGAG 1395
319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyr 338
1396 GAGTACCGCAAGTACGTGAGAGAGGACGACGCTTCAGCGCGGTTCAGCAGGTGTTTC 1455
339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyr 358
1456 GTCGCGAGCGGAGCTGCCGACACCGTCAGCATCTCAGCGGCTCAAGGAGAGTAC 1515
359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeuSer 378
1516 GAGGGCACCATGGCTGAGGATCCAGGACCGCGCTCTGTCGCGCACAGCTCTCC 1575
379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1576 GCGAGTACATCATGCTGCGCACCTGCTGACAAAGCCATAGACTGTTGACGAGGCC 1635
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
1636 TGGCCCAATCTGAGGTGAGCTCGACAGCCAGCGGAGAGATTGATAACCTCGAGAGG 1695
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1696 AAGAGATCCAGTTCAGGTGAGTCCACGCGCTCGAAGAGAGAGGACAAGCCAGC 1755
439 LysAlaArgLeuIleGluVal----- 445
1756 AAGCCCGGTGATTGAGGT-GGCTGCTCGAGCACTGAATTTCTCAAACAAGTCTCTC 1814
445 ----- 445
1815 TTGCTCATGTTCTCTGTTTGAATTTTGAATACTAACGCTGCTTAACGTGCGCATTTGT 1874
446 -----ArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyr 463
1875 CAGGTCAAGAGAAATTGACGATCTGAGGACAAAGCTCGACCCCTGACCATGAGTAC 1934
464 ArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeu 483
1935 AGGAAGAGAGAGAGAGAAATTGATGATGATGAGAGAGCTGAAGCAGCGCGCGAGGCTC 1994

QY 484 MetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAlaAspLeuArg 503
Db 1995 CAGTTCACCTTCGAGGAGCGGAGCGCGGATGAGCCTGGCCCTGTGGCGGACCTCAAG 2054
QY 504 TyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGlu 523
Db 2055 TACGGCGCTTCAGGAAATCGACGCTGCTATCTCCAAAGCTGGAG---AGCGAAACAGGG 2111
QY 524 GluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu----- 539
Db 2112 GAGAACCTGATGCTCCGCAACCGCTCGGCCCTGAACAAATTCAGAGGTATGTTATTAT 2171
QY 539 ----- 539
Db 2172 TCTTGTTCACCGTCACAAATAATTTGCGAGCAAGTCCGAAATTTGCGGATCGTCTCTTA 2231
QY 540 -----ValValSerAr 543
Db 2232 GTAGAGTAGTCGTGCGCGTCTGAAATGGTGTTCGTCTATGGCGAGGTGTGTAGCCG 2291
QY 543 GTrpThrGlyIleProValThrArgLeuGlyGlnAsnGlnLysGluArgLeuIleGlyLe 563
Db 2292 TTGACGCGGTATTCAGTGACCGCGCTTGGCCAGCAACAGAGAGAGGTGTGTGGCCT 2351
QY 563 uAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGl 583
Db 2352 GGCTGACAGCGCTTCACAGAGGGTGTTCGCGCAGACAGAGGCTGTGAGCGCCCTCGCAGA 2411
QY 583 uAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLe 603
Db 2412 GCGCGCTGAGGTGCGAGGCGCGCTTTCGCGAGCCACAAACAGCCCACTGGCTCGTTCT 2471
QY 603 uPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLe 623
Db 2472 CTTCTGGTCCGACTGGCGTGGGAAAACCTGAGTGTGCGCAAGGCGCTAGCCCAACAGCT 2531
QY 623 uPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSe 643
Db 2532 GTTCGACGACGAGAACCTCTTGTTCGCGCATCGACATGTCGAGTACATGCGAGCAGCACTC 2591
QY 643 rValSerArgLeuIleGlyAlaProGly----- 653
Db 2592 GGTTCGCGCGCTCATCGAGCAGCACCTCGG-GTAAAGTAGCAGAAATGCAATGTCATCTTCG 2650
QY 653 ----- 653
Db 2651 TATTTTAACTGCGAATTGCGATGAATCTTGTTCGACAGTGGCGCGGACCTGTGACAG 2710
QY 654 -TyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgArgProTyrCy 673
Db 2711 CTACGTGGCCCATGAAGAGGTGGCGAGCTGACTGAACAAGTGGAGGAGGCGGTACAG 2770
QY 673 sValIleLeuPheAspGluValGluLysAlaHisAlaValAlaPheAsnThrLeuLeuGl 693
Db 2771 CGTGATCTGTTCACGAGTCCAGAGAGGCGCCATGTGCGCGTGTTCACACACCTGTCTCA 2830
QY 693 nValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSe 713
Db 2831 GGTCTCTGACGACGCGAGGTGACGATGGGCAAGGCGAGCGGTGACCTTCAGGAACAC 2890
QY 713 rValIleLeuMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLy 733
Db 2891 CGTGATCATGACATCGAACTCGCGCGCGGAGCACCTCTCTGCTGGGATGTGTGGGCAA 2950
QY 733 sValThrMetGluValAlaArgAspCysValMet-Arg----- 745
Db 2951 GAACTCCATGAAGTGTGCTCGCATCTGGTTCATGCGAGGAGGTATGCAATTCGATAACGTT 3010
QY 745 ----- 745
Db 3011 CTGNACTGCTGTAGACATGTTTCTCTCGCAGCGGTGATGCTCAATCGGATCGTTATTGCT 3070

746 -----GluValArgLysHisSheArgProGluLeuLeuAsnArgLeuAspGluLeuValV 764
 3071 GTGTGAGTGGAGGACCTCCGCCCTGAGTCTGAACCGTCTGACGAGATCGTGA 3130
 764 alPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspV 784
 3131 TCTTCGATCTCTGTCCACGAGAGCTGAGAGGTGCTGCGCTTCAGATGAAGATG 3190
 784 alAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspT 804
 3191 TGGCGTCCGTCTTCCGAAAGGGGCATCGCTGCTGTGACCGACGCCGATGACA 3250
 804 yrlleLeuAlaGluSerThrAspPro----- 812
 3251 TCATCTTGTCTCTCTTACGATCCGGTATGTGACCAATCCATGATTTGATCCATCTGAAT 3310
 812 ----- 812
 3311 TCGTCGGTGACCTGATGGTGTGACTCTCTTATCTTTTGTGTGGCTTCAACAACAG 3370
 813 ValTyrglyAlaArgProIleArgArgTrpMetGluLysLysValValThrGluLeuSer 832
 3371 GTGTATGGCGCGGCAATCAGCAGGTGGATCGAGAGAGGTGGTGACGACTCTCG 3430
 833 LysMetValValArgGluGluLeuAspGluAsnSerThrValTyrlleAspAlaGlyAla 852
 3431 AAGATGCTGATCCAGGAGGAGATCGACGAGAACTGACGGTCTACATCGACGCCGCC 3490
 853 Gly-----AspLeuValTyArgValGlu---SerGlyGlyLeuValAspAlaSerThr 869
 3491 GGCAAGCAGCAGCTGTCTACAGGTGACCGGAGCGCGGTCTGTGTAACGCTGAGACG 3550
 870 GlyLysLysSerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAla 889
 3551 GGGATGAAGTCGACATCTTGATCCAGTCCCAACAGCTCCACACGAGCGCTGCG 3610
 890 GlnAlaValLysLysMetArgIle---GluGluLeuAspAspAspAenGluGlu 907
 3611 CAGCCCTCAAGAAGATGAGGATCATCGAGGAGGACGAGGACCGCATGGAGGAG 3667

JUL 12
 AAI66071
 AAI66071 standard; DNA; 2045 BP.
 AAI66071;
 14-JAN-2002 (first entry)
 Zea mays 101 kDa heat shock protein gene (GenBank: AF083327).
 Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana; ds.
 Zea mays.
 W0200170929-A2.
 27-SEP-2001.
 20-MAR-2001; 2001WO-US08836.
 20-MAR-2000; 2000US-190769P.
 18-APR-2000; 2000US-198116P.
 (ARCH-) ARCH DEV CORP.
 Lindquist S, Queitsch C, Vierling E;
 WPI; 2001-639123/73.
 P-PSDB; AAM51670.

Transgenic plants with improved heat stress tolerance, useful for

PT producing animal feed, oil and synthetic products -
 XX Claim 4; Page -; 91pp; English.
 XX

CC The invention relates to a transgenic plant, comprising a genetic
 CC construct comprising a promoter operatively linked to a nucleic acid
 CC sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 CC increased stress tolerance, especially to heat. The plant is a cereal,
 CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 CC synthetic product-producing plant, an environmental waste absorbing
 CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 CC an animal feed plant. In particular, the transgenic plant is cotton,
 CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 CC Arabidopsis thaliana. The plants may be used to produce animal feed,
 CC alcohol, crop, oil, medicine or a synthetic product.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.

XX
 SQ Sequence 2045 BP; 470 A; 532 C; 673 G; 370 T; 0 other;

Alignment Scores:

Pred. No.:	2,52e-176	Length:	2045
Score:	2502.50	Matches:	497
Percent Similarity:	93.65%	Conservative:	49
Best Local Similarity:	85.25%	Mismatches:	32
Query Match:	55.00%	Indels:	5
DB:	22	Gaps:	4

US-09-812-350-17 (1-911) x AAI66071 (1-2045)

QY	329	AlaPheGluArgArgPheGlnGlnValTyValAlaGluProSerValProAspThrIle	348
DB	3	GGTTCGAGCGCGGTTCAGCAGGTGTTTCGCGGAGCGAGCGTCCCGATACCGTC	62
QY	349	SerIleLeuArgGlyLeuLysGluLysTyrgluGlyHisHisGlyValArgIleGlnAsp	368
DB	63	AGCATTTCTGAGGGGACTCAAGGAGAAGTACGAGGGGCCACCATGGCGTGAGGATCCAGAC	122
QY	369	ArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArgTyrlleThrGlyArgHisLeuPro	388
DB	123	CGCGCCCTCGTGTGTCGCGGCACAGCTCTCCGCGAGGTACATCATGGTGGCGCACCTGGCT	182
QY	389	AspLysAlaIleAspLeuValAspGluAlaCysAlaAenValArgValGlnLeuAspSer	408
DB	183	GACAAAGCCATAGACCTGGTGGACGAGGCTTCGCGCAATGAGGGTGCAGCTCGACAGC	242
QY	409	GlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHis	428
DB	243	CAGCCGAGGAGATTGATAACTCGAGAGAGAGAGATCCAGCTTGAGTGGAGTCCAC	302
QY	429	AlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGlu	448
DB	303	GGCTCTGAGAGGAGAGAGACAGCCAGTAAAGCCCGCTGATTGAGGTTCAGGAAGGAA	362
QY	449	LeuAspAspLeuArgAspLysGlnProLeuThrMetLysTyrgluLysGluLysGlu	468
DB	363	TTGGACGATCTGAGGACACAGCTGCGAGCCCTTGACCATGAAGTACAGGAGAGAGAG	422
QY	469	ArgIleAspGluIleArgArgLysGlnLysArgGluGluLeuMetPheSerLeuGln	488
DB	423	AGAATCGATGATGATCAGGAAGCTGAAGCAGCGCGCGAGGAGTCCAGTTCACCTCGAC	482
QY	489	GluAlaGluArgArgTyrgluAlaAlaArgAlaAlaAspLeuArgTyrglyAlaIleGln	508
DB	483	GAGCCGAGCGCGGATGACCTGCGCGGTGTGGCCGATCTCAAGTACGCGCGCTCCAG	542
QY	509	GluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAenValMetLeu	528
DB	543	GAATCGAGCGCGGATCTCCAGCTGAG---AGCGAAACAGCGGAGAGACCTGATGCTC	599
QY	529	ThrGluAenValGlyProGluHisIleAlaGluValValSerArgTrpThrGlyIlePro	548

600 ACCGAAACCGTCGCCCTGACAAATTCGAGAGGTGTCGAGCGGTATTTCCA 659
549 ValThrArgLeuGlyGlnAsnGluLeuGluArgLeuLeuLeuLeuHis 568
660 GTGACCCCGCTGTGCGCCAGACGAGAGGCTGTGTCGCGCTGCTGACAGGCTTCAC 719
569 LysArgValValGlyGlnAsnGlnAlaValSerGluAlaLeuLeuArgSer 588
720 CAGAGGTGTGTCGCCAGACAGAGGCTGTGAGCGCTGTCAGAGGCGGTGCTGAGTGC 779
589 ArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuGlyProThr 608
780 AGGCGCGCTGTGCGCAGCCACACAGCCACTGCTGCTCTCTCTCTCTCTCTCT 839
609 GlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspGluAsn 628
840 GCGCTGGGAAACTGAGCTGGCCAGCCCTAGCCGACAGCTGTTCGACGAGAAC 899
629 LeuLeuValArgIleAspMetSerGlyThrMetGluGlnHisSerValSerArgLeuIle 648
900 CTTCTGTGCGCATCGACATGTCGAGTACATGAGCAGCAGCTCGGTGCTCGCTCATC 959
649 GlyAlaProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 668
960 GGAGCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
669 ArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPhe 688
1020 AGGAGCGGTACAGCGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079
689 AsnThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrVal 708
1080 AACACCTGTCTCCAGTCTCTGAGCAGCGGCTGACGAGTGGGAGGAGGAGGAGGAGG 1139
709 AspPheArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAla 728
1140 GACTTCAGGAACACCGTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1199
729 GlyLeuThrGlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArg 748
1200 GGGATGTGGGCAAGACTCCATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
749 LysHisPheArgProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeu 768
1260 AGGCACCTTCGCGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1319
769 SerHisAspGlnLeuArgLysValAlaArgLeuGluMetLysAspValAlaValArgLeu 788
1320 TCCACAGCAGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1379
789 AlaGluArgGlyValAlaLeuAlaValThrAspAlaLeuAspTyrIleLeuAlaGlu 808
1380 GCCGAAAGGGCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1439
809 SerTyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysValVal 828
1440 TCTTACATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1499
829 ThrGluLeuSerLysMetValValArgGluGluLeuAspGluAsnSerThrValTyrIle 848
1500 ACGAGTGTGAGAGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1559
849 AspAlaGlyAlaGly-----AspLeuValTyrArgValGlu----SerGlyGlyLeuVal 865
1560 GAGCGCGCGCGGTGAGGAGCAACTGCTACAGGTGAGCGGAGCGCGCTGCTGCTGCTGCT 1619
866 AspAlaSerThrGlyLysLysSerAspValLeuLeuHisIleAlaAsnGlyProLysArg 885
1620 AAGCTGAGACGGGATGAGTGGACATCTCTGATCCAGTCCCGCCACCGCTCCACGAG 1679
886 SerAspAlaAlaGlnAlaValLysLysMetArgIle---GluGluLeuGluAspAsp 904

Db 1680 AGCGACGCTGCGAGCCGTCAGAGATGAGATCATGAGGAGGACGAGCGCATG 1739
QY 905 AsnGluGlu 907
Db 1740 GACGAGGAG 1748
RESULT 13
AAI66072
ID AAI66072 standard; DNA; 2045 BP.
XX AAI66072;
XX 14-JAN-2002 (first entry)
XX Zea mays heat shock protein 101 gene (GenBank: AF077337).
XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
XX canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX Arabidopsis thaliana; da.
XX Zea mays.
XX WO200170929-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-US08836.
XX 20-MAR-2000; 2000US-190769P.
XX 18-APR-2000; 2000US-198116P.
XX (ARCH-) ARCH DEV CORP.
XX Lindquist S, Queitsch C, Vierling B;
XX WPI; 2001-639123/73.
XX P-PSDB; AAM51663.
XX Transgenic plants with improved heat stress tolerance, useful for
XX producing animal feed, oil and synthetic products -
XX Claim 4; Page -; 91pp; English.

CC The invention relates to a transgenic plant, comprising a genetic
CC construct comprising a promoter operatively linked to a nucleic acid
CC sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
CC increased stress tolerance, especially to heat. The plant is a cereal,
CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a
CC synthetic product-producing plant, an environmental waste absorbing
CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or
CC an animal feed plant. In particular, the transgenic plant is cotton,
CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
CC Arabidopsis thaliana. The plants may be used to produce animal feed,
CC alcohol, crop, oil, medicine or a synthetic product.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained from GenBank using the Accession Number
CC reference provided in the specification.

SQ Sequence 2045 BP; 470 A; 532 C; 673 G; 370 T; 0 other;

Alignment Scores:

Pred. No.:	2,52e-176	Length:	2045
Score:	2502.50	Matches:	497
Percent Similarity:	93.65%	Conservative:	49
Best Local Similarity:	85.25%	Mismatches:	32
Query Match:	55.00%	Indels:	5
DB:	22	Gaps:	4

US-09-812-350-17 (1-911) x AAI66072 (1-2045)

QY 329 AlapheGluArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIle 348
|||||

3 GCGTTCGAGCGCGGTTCACGAGGTGTTCTCGCGGAGCGCGGTGCCGATACCGTC 62
 349 SerLeuArgGlyLeuLysGlyTyrGluGlyHisHisGlyValArgIleGluAsp 368
 63 AGCATTCGAGGGGACTCAGAGGAGTACGAGGGGACCATGGCGTGAGGATCCAGAC 122
 369 ArgAlaLeuIleAsnAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuPro 388
 123 CGCGCCCTCGTGGCGGCGACAGCTCTCCGCGAGGTACATCATGGGTGCGCACTGCT 182
 389 AspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSer 408
 193 GACAAAGCCATAGACCTCGTGGAGCAGGCGCTCGCCCAATGTGAGGGTGCAGCTCGACAGC 242
 409 GlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHis 428
 243 CAGCCGAGAGGATTGATTAACCTGGAGAGAGAGATCCAGCTTGAGGTCCAGCTCCAC 302
 429 AlaLeuGluArgGlyLysAlaSerLysAlaArgLeuIleGluValArgLysGlu 448
 303 GCGCTCGAGAGAGAGAGCAAGGCCAGTAAAGCCCGGCTGATTGAGTCCAGGAAGAA 362
 449 LeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGlu 468
 363 TTGGACGATCTGAGGAGCAAGCTGAGCGCCCTGACCATGAAGTACAGGAAGGAGAG 422
 469 ArgIleAspGluIleArgArgLysGlnLysArgGluGluLeuMetPheSerLeuGln 488
 423 AGAATCGATGATCAGAGAGCTGAGAGCGCGCGAGAGGCTCCAGTTCCACCTGCAG 482
 489 GluAlaGluArgArgTyrAspLeuAlaArgAlaAspLeuArgTyrGlyAlaIleGln 508
 483 GAGCGGAGCGCGGATGACCTGCGCCGCTGTGGCGGATCTCAAGTACGCGCGCTCCAG 542
 509 GluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeu 528
 543 GAAATCGAGCGCGGATCTCAAGGCTGGAG---AGCGAAACAGGGGAGAGCTGATGCTC 599
 529 ThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIlePro 548
 600 ACCGAAACCGCTCGGCCCTGAACAAATTCAGAGGTGGTGAGCCGTTGACCGGGTATCCA 659
 549 ValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHis 568
 660 GTGACCCGCTTGCCGACAGCAACAGAGAGGCTGGTGGCGCTGGCTGACAGGCTTCC 719
 569 LysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSer 588
 720 CAGAGGGTGGTGGCCAGACAGAGGCTGTGAGCGCGCTCGCAGAGCGCGGTGCTGAGGTG 779
 589 ArgAlaGlyLeuGlyArgAlaGlnProThrGlySerPheLeuPheLeuGlyProThr 608
 780 AGGGCGGTCTTGGCAGGCCAACAGACCCACTGGCTGTTCTCTTCTTCTGGTCCGACT 839
 609 GlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsn 628
 840 GCGGTGGGGAACCTGAGCTGGCCCAAGGCCCTAGCCGCAACAGCTGTTTCAGACGAGAAC 899
 629 LeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIle 648
 900 CTTCTGTGCGCATCGATGCGGAGTACATGGAGCAGCACTCGGTGCTCGGCTCATC 959
 649 GlyAlaProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 668
 960 GGAGCACCACTGGTACGTCGCGCCATGAAGAGGGTGGGAGCTGACTGAACAAGTCAGG 1019
 669 ArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPhe 688
 1020 AGGAGGCGGTACAGCGTATCTGTTTCAGAGGTCGAGAAAGGCCCTGTCGCGCTGTTTC 1079
 689 AsnThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrVal 708
 1080 AACACCCCTGCTCCAGTCTCTGACAGCGAGGCTGACGGATGGGCAAGGAGGAGCGGTG 1139

709 AspPheArgAsnSerValIleIleMetThrSerAsnLeuGluValAlaGluHisLeuLeuAla 728
 1140 GACTTCAGAAACACCGGTGATCATCATGATCATGACATCGAACCTGGCGCCGAGCACCTCTCGT 1199
 729 GlyLeuThrGlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArg 748
 1200 GGGATGTTGGGCAAGAACTCCATGAAGTTCGCTCGCATCTGGTCTATGAGGAGGTGAG 1259
 749 LysHisPheArgProGluLeuAsnArgLeuAspGluIleValValPheAspProLeu 768
 1260 AGGCATCTCGCCCTGAGCTGCTGAACCGTCTCGACGAGATCGTATCTTCGATCTCTCTG 1319
 769 SerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeu 788
 1320 TCCACGAGCAGCTGAGGAGGTGCTGCTCGCTTCAGATGAGGATGTGGCCGCTCTCTT 1379
 789 AlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGlu 808
 1380 GCCGAAAGGGGATCTCTGCTGTGACCGCAGCGCATTCGACATCATCTTGTCTCTC 1439
 809 SerTyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValVal 828
 1440 TCTTACGATCCGCTGTACGCGCGCGCCCAATAAGAGGTGGATCGAGAAGAGGGTGGTG 1499
 829 ThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIle 848
 1500 ACCAGCTGTGAGAGATGCTGATCCAGGAGGAGATCGACGAGACTGACCGCTTACATC 1559
 849 AspAlaGlyAlaGly-----AspLeuValTyrArgValGlu---SerGlyGlyLeuVal 865
 1560 GACGCGCGCGCGGTAAAGGACGAACCTGGTCTACAGGCTGACCGAGCGCGGTCTGGTG 1619
 866 AspAlaSerThrGlyLysSerAspValLeuIleHisIleAlaAsnGlyProLysArg 885
 1620 AACGCTGAGACGGGGATGAAGTCGACATCTCGATCCAGTCCCCACAGCTCCACAGG 1679
 886 SerAspAlaAlaGlnAlaValLysLysMetArgIle---GluGluIleGluAspAsp 904
 1680 AGCAGCGCTCGCAGCGCTCAAGAAGATGAGGATCATGAGGAGGAGGAGGAGCGCATG 1739
 905 AsnGluGlu 907
 1740 GACGAGGAG 1748
 RESULT 14
 AAI66083
 ID AAI66083 standard; DNA; 7726 BP.
 AC AAI66083;
 XX
 XX 14-JAN-2002 (first entry)
 XX
 XX Leishmania donovani heat shock protein 100 gene (GenBank: Z94053).
 DE Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana; ds.
 XX
 OS Leishmania donovani.
 XX
 PN WO200170929-A2.
 PD 27-SEP-2001.
 XX
 PF 20-MAR-2001; 2001WO-US08836.
 XX
 PR 20-MAR-2000; 2000US-190769P.
 PR 18-APR-2000; 2000US-198116P.
 XX
 PA (ARCH-) ARCH DEV CORP.
 XX
 PI Lindquist S, Queitsch C, Vierling E;

Qy	181	VallelGlyArgAspGluGluLeuIleArgValValValArgIleLeuSerArgThrLys	200
Db	2980	GTATTATGGTCGCAGCAGAGATTCTCGGCACATTCTCGTGTGTGTGCACGTCCACCAAG	3039
Qy	201	AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu	220
Db	3040	AACAAACCCGGTGTGATCGGGAGACTGGAGTGGGTAAAGACCCGGATTGTGGAAGGCAAT	3099
Qy	221	AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer	240
Db	3100	GCACAGCAGGTGGTCGAGCGCAGCTGCCGCACACTCTCTCAGGTATTGCATCTTCTCA	3159
Qy	241	LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluArgLeu	260
Db	3160	CTGCATCATGGCCGCTGATCGCCGCCCAAGTACCGCGCGAGTTCGAGGAGCGCGCTG	3219
Qy	261	LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu	280
Db	3220	AAGCCGCTGTGAATGAGGTGAAGGAAAGCGACAATAAGATCATCTCTTTCATTGACGAG	3279
Qy	281	IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe	300
Db	3280	ATCCACTCTGTCTCGCGCCGCAAGTCCGACGGCGCATGACGCCGCGAATCTGCTG	3339
Qy	301	LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr	320
Db	3340	AAGCCGCTGTGCCCGTGGCGAGCTGGCACGATCGTGTCCACACGCTCGAGGAGTAC	3399
Qy	321	ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrValAla	340
Db	3400	CGCCCATGCTGGAGAAGACGCCGCTTCGACGCCGCTTCATGTCAGTGCAGGTGAAT	3459
Qy	341	GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly	360
Db	3460	GAGCCATCTGTCCGAGGTGTACAGCATCTCTCGTGTGTCTGAAAGACCGCTACGAGCAG	3519
Qy	361	HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg	380
Db	3520	CACCACGGTGTGAGATCAGCAAGAGCGGTGTGTGTGGCGCGCATGTGGCTGCGCCG	3579
Qy	381	TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla	400
Db	3580	TACATTACGAACCGCTTCTGCCGGAATAGGCTATTGACCTTATTGATGAAGCTGCGCC	3539
Qy	401	AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg	420
Db	3640	AACGTGCGGTGACGTTGTCTGTCGCGCGCGCGCAAAATCGACGCCCTCGAGCGCAAGAAG	3699
Qy	421	MetGlnLeuCluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla	440
Db	3700	CGCCACTGGAGATCAGGAGAAAGCGCTGCAGCGCGCAAGGACCGCATCGGCCAAGGAG	3759
Qy	441	ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr	460
Db	3760	CGGCTGAAAGCCGTGAAGCGGAGATTCAAGAGGTGGAGGAGAAGCTTGGCCCATCTCTT	3819
Qy	461	MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg	480
Db	3820	GCCAAAGTACGACAGAGCTGGCCGTATCGACGAGCTCGACGCAACCGCAAGCAAGCTG	3879
Qy	481	GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAla	500
Db	3880	GACGAGAAGAGGTGAAGCTAGACGGCGGGAGCGGATCGGTGACATGAGAGCGGACGA	3939
Qy	501	AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu---GluGly	519
Db	3940	GATCTTTAAATACAAATGTCATCCGATCTCCAGGACAGGATCCGATCCCTCAAGGAGGAG	3999
Qy	520	ThrSerSerGluCluAsnValMetIleThrGluAsnValGlyProGluHisIleAlaGlu	539
Db	4000	ATCGAAGACGACGACGACGATGCTGCAATGCGACCGTCACTGAGACGCGACATCGCCACC	4059
Qy	540	ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg	559

4060 GTTGTGCTGGAGCATATCCCGGTGACAAAGCTGACCGACCGACCGAGCCG 4119
 560 LeuIleGlyLeuAlaLeuArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
 4120 CTGCTGCACCTGGCCGACCACTGCACTCGGTGAAAGGCCAGGATGAGCGGTGAGC 4179
 580 AlaValSerGluAlaLeuArgSerArgAlaGlyLeuGlyGlnGlnProThr 599
 4180 CGTGTGCGGAGGCTATCTGCGCTCAGTGGCGGCTGCGCCGCTGCGACCGCCACC 4239
 600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
 4240 GGCTCTTCTGCTTCTGCGCCCGCCCGCGGTGGCAAGACAGAGCTGTCCAGGCGCTC 4299
 620 AlaGluGlnLeuPheLeuAspGluAsnLeuLeuValArgLysLeuMetSerGluTyrMet 639
 4300 GCTCAAGAGCTCTTCGACGATGCAAGATCATGTGCGATGCGACATGAGCGAGTACATG 4359
 640 GluGlnHisSerValSerArgLeuLeuGlyValAlaProProGlyTyrValGlyHisGluGlu 659
 4360 GAGCAGCATTCTGTGGCGGCTGATCGCGCGCCCGCGGCTGCTGCTGCTGCTGATGAG 4419
 660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGlu 679
 4420 GGTGGCCAGCTGACGAGCGGTGGCGGTGCGCCATACACGCTGCTGCTGCTGATGAG 4479
 680 ValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArg 699
 4480 GTGGAGAGCGCCACCGACGCTTCAAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4539
 700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSer 719
 4540 TTGACTGACTCGACGCGCGCAACCGTGAATTTCTGCTACGATCATCATGATCATCC 4599
 720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
 4600 AACTGGTGGCGAGTATCTGCAAGACATGACCTCACCGAAGCGCTACGAGTGGCA 4659
 740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu 759
 4660 CAGACACAGGTGATGGCGAGGTGAGGAAATTTCTTTCGCGCGAGTTCATCAACCGACTG 4719
 760 AspGluLeuValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
 4720 GATGACATCATCTCTTCGCTCTCTGGCTTAAAGAGATGACCGGATCATCGACCTC 4779
 780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
 4780 ATCACCAGAGAGCTCTATGGCCGCTCAAGGACAGCTCCATCGCGGTTCCTCCACAGAA 4839
 800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyLysAlaArgProIle 819
 4840 GAGGCTAAGAACTACGATGATGGAGTGGCTTCGACCGGATATGGCGCTCGCCACATA 4899
 820 ArgArgTyrMetGluLysValValThrGluLeuSerLysMetValValArgGluGlu 839
 4900 CGCGCTGGGTTCAGAGAACATCACACAGAGCTCAGCGCGATGATCATCTTCGCAAGAG 4959
 840 IleAspGluAsnSerThrValTyrIle 848
 4960 CTGTGCGCGCAACAGCAGCGGTGAAGTG 4986

SU15
 166084

AAI66084 standard; DNA; 5997 BP.

AAI66084;

14-JAN-2002 (first entry)

Leishmania major heat shock protein Hsp100 gene (GenBank: Z38058).

KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 XX Arabidopsis thaliana; ds.
 OS Leishmania major.
 XX WO200170929-A2.
 XX 27-SEP-2001.
 XX 20-MAR-2001; 2001WO-US08836.
 XX 20-MAR-2000; 2000US-190769P.
 XX 18-APR-2000; 2000US-198116P.
 XX (ARCH-) ARCH DEV CORP.
 XX Lindquist S, Quesitsch C, Vierling B;
 WPI: 2001-639123/73.
 DR P-PSDB; AAM51668.
 XX Transgenic plants with improved heat stress tolerance, useful for
 PT producing animal feed, oil and synthetic products -
 XX Claim 4; Page -; 91pp; English.
 XX The invention relates to a transgenic plant, comprising a genetic
 CC construct comprising a promoter operatively linked to a nucleic acid
 CC sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 CC increased stress tolerance, especially to heat. The plant is a cereal,
 CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 CC synthetic product-producing plant, an environmental waste absorbing
 CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 CC an animal feed plant. In particular, the transgenic plant is cotton,
 CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 CC Arabidopsis thaliana. The plants may be used to produce animal feed,
 CC alcohol, crop, oil, medicine or a synthetic product.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.
 XX SQ Sequence 5997 BP; 1182 A; 1830 C; 1645 G; 1340 T; 0 other;
 Alignment Scores:
 Pred. No.: 6,46e-160 Length: 5997
 Score: 2289.00 Matches: 461
 Percent Similarity: 72.08% Conservative: 151
 Best Local Similarity: 54.30% Mismatches: 227
 Query Match: 50.31% Indels: 10
 DB: 22 Gaps: 5
 US-09-812-350-17 (1-911) x AAI66084 (1-5997)
 QY 3 ProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAlaVal 22
 Db 2187 CCAGAA---TGAAGACGAGCGCGCTGCGACTTGATGCTGCGATGCGGCGACTGGCAGC 2243
 QY 23 AsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLysSerAspPro 42
 Db 2244 AAGAAGGCAACCGCTACCTCGACCCCGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2303
 QY 43 ThrGlyIlePheProGlnAlaLysSerSerAlaGlyGlyGluAsnAlaAlaGlnSerAla 62
 Db 2304 AACAGTCTTCTCCCGTGGCGCGCAAACTCGT-----GCCGCTCAGTC 2351
 QY 63 GluArgValIleAsnGlnAlaLeuLysLysLysLysLysLysLysLysLysLysLysLys 82
 Db 2352 AAGACGAGCTGGAGGCGCGCTGTCGACGCGATTCACACGAGTGTGCGCCACGCGAG 2411
 QY 83 IleProLysSerSerSerLeuIleLysValIleLysValIleLysValIleLysValIle 102
 Db 83 IleProLysSerSerSerLeuIleLysValIleLysValIleLysValIleLysValIle 102

2412 CGCGGGCCCACTCGGACATGATCGTGTCTCATGAACCGCGGAGCAGGAGCGTGTGCC 2471
103 ArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSerGln 122
2472 CTCGGTGACACCTCATCGCGCTGACACCTTCTCTGGCCCTGCGACGAGCAAGGAG 2531
123 IleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGluValGlu 142
2532 GTTGAAGAATCTGGATGCTGCGAGGGCTGGAAGAAGGCGATTCTGCGCCACGCTGCT 2591
143 LysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsnPheGln 162
2592 GAGATGCC-----AAGGGAAAGAAAGATACCTCCGACTTCCAAAGACGACAACTACGAG 2645
163 AlaLeuLysThrThrGlyArgAspLeuValGluGlnAla-----GlyLysLeuAspPro 180
2646 TCGCTGAACAGTACGCGAGTTGATCTCTGCAAGCAGCGGAGGAGGAGAAAGCTTGGACCCG 2705
181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
2706 GTTATTGGTCGCGCAGACGAGATTCTGCGCACTATCCGCGTGTCTCACGCCGACCAAG 2765
201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
2766 AACAAACCGGTGCTTATCGGGAGCCTGGAGTGGGTAAAGACCGCGATTGTGGAGGGCATT 2825
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
2826 CGCGACAGGTGGTGGCGAGCGAGTCCGCGACACCTCTCCGGTATTCCATCTCTCA 2885
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGluGluArgLeu 260
2886 CTGGACATGGCGCGCTGATCGCGCGTGGCAAGTACCGCGCGAGTTCGAGGAGCGCTG 2945
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
2946 AAGCCCTGTGTAATGAGGTGAAGGAAGCGCAATAAGATCAATCTCTTTATGTGCGAG 3005
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
3006 ATCCACCTGTGTCGCGCGCGGCAAGTCCGCGCGCGATGACGCGCGGATCTGCTG 3065
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlyTyr 320
3066 AAGCCGTGTGTCGCGCGTGGCGAGCTGCGCAGCATCGGTGCCCAACCGTTCGAGGAGTAC 3125
321 ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrValAla 340
3126 CGCCAGTACCTGAGAGAGAGCGCGCTTCGAGCGCGCTTCATGCCAGTGCAGGTGCAT 3185
341 GluProSerValProAspThrIleSerIleLeuArgLysGluLysTyrGluGly 360
3186 GAGCCGTGACGTGAGGAGTGCAGGATCCTCGCTGGTCTGAAGACCGCTACGAGCAG 3245
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeuSerAlaArg 380
3246 CACCACGTGTGAGATCAGGACAGCGCGTGTGTCGCGCGCGAGTGTGCTGCGCGC 3305
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
3306 TACATTACGAACCGCTCTCTGCGCGATAGGCGATTGACCTCATATTATGAGCGTGGCC 3365
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
3366 AAGTGGCGCTGACGTGTGTCGCGCGCGCGGCGAAATCGACGCCCTCGAGCGCAAGAG 3425
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
3426 CGCCAGTGGAGATCGAGGAGAGGGCAGCTGACGCGGCAAGAGACGATCGGCCAAGGAG 3485
441 ArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysAlaLeuGlnProLeuThr 460
3486 CGGCTGAAGCCGTGAAGCGGAGATTTCAGAGGTGAGGAGGAACTTGGCCCTCTACTT 3545

461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
3546 GCCAAGTACAAAGAGGCGTGGCCGTATTGACGAGCTGACGCAACGCGCAAAAGCTG 3605
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
3606 GACGAGAAGAGGTGAGCTGGAGCGGCGAGCGGATGCGGACATGGAGACGCGACGA 3665
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu---GluGly 519
3666 GACCTCAATACCGGTCTATCCCAATCTCCAGGACAGATCCGATCCCTCAAGGAGCG 3725
520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
3726 ATCGAGACGACGAGCGGAGCGATGCTGCGAGGCGACCGTGACCGGACGCGACATCGCCACC 3785
540 ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg 559
3786 GTTGTGTGCGCTGCGACCAATATCCGCTGACAAAGCTGAGCCAGACCGAAGCGAGCGC 3845
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
3846 CTGTTGCACTTCTGCGCGACGCTGCCTCGTGTGAAGGCGCAGGATGAGCGCGTGCAG 3905
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
3906 CGTGTGCGGAGCGCATCTGCGCTCACGTGCGCGCTGGCCGCTTGGAGCCGCGCCACC 3965
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
3966 GCGCTTTCTCTGTTCTCGTCCACCGCGCTGGGCAAGACAGAGCTATCCAAAGCGCGTT 4025
620 AlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMet 639
4026 GCGTCAGAGCTCTTCGACGACGCCAAGTACATGTTGCGACTTGACATGAGCGAGTACATG 4085
640 GluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGlu 659
4086 GAGCAGCACTCTGTGGCGCGCTGATCGGACCCCGCGGGGTACGTCGCCCAAGAA 4145
660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGlu 679
4146 GGTGCCACAGCTGACGAGCGCGTGGCGCTGCCATACACGCTGGTGTCTGATGAA 4205
680 ValGluLysAlaHisValAlaValPheAsnThrLeuGlnValLeuAspAspGlyArg 699
4206 GTGGAAGGCGCACCCGACGCTCTCAACGTACTGTGCGAGGTGCTCGACGAGCGCGG 4265
700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSer 719
4266 TTGACTGACTCGACGCGCGCAACCGTGAATTTTCGCAACAGCATCATCATGACATCC 4325
720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
4326 AACTTGGTGGCGAGTATCTGCAGAACATGCAGACCTTCACGGAAGCGGTACGAAGTGGCA 4385
740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu 759
4386 CAGGCACAGGTGATGGCGGAGGTAAAGAAAGTTCTTTCGCGCGAGTTTCATCAACCGACTG 4445
760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
4446 GACGACATCATCTCTTCCTCGCTCTCTGGGCTTAAAGAGATGACCGCATCATCGACCTC 4505
780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
4506 ATCAGAGAGGCTCAATGGCGCGCTCAAGGACAGTCCATCGGGGTCTCCCTCACCGAA 4565
800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIle 819
4566 GAGGCCAAGCAGTACGTATTGGAGTTCGCGCTTCGAGCGGATATGGCGCGCTCGCCGCTA 4625

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 12, 2004, 22:27:44 ; Search time 72 Seconds
(without alignments)
2008.332 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

quence:

1 MNPEKFTKHTNETATHEL.....VKMKRIEIEDDNEEMTED 911

ring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

urched: 1107863 seqs, 158726573 residues

al number of hits satisfying chosen parameters: 1107863

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase : A_Geneseq_13Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	Length	ID	Description
1	4550	100.0	911	22	AA151659 Arabidopsis thalia
2	4550	100.0	911	22	AA151660 Triticum aestivum
3	4545	99.9	911	22	ABP81189 Arabidopsis thalia
4	4003	88.0	911	22	AA151662 Glycine max heat sh
5	3896	85.6	909	22	AA151665 Nicotiana tobacum
6	3883.5	85.4	912	22	AA151663 Zea mays heat choc
7	3883.5	85.4	912	22	AA151669 Zea mays heat choc
8	3822.5	84.0	913	22	AA151664 Triticum aestivum
9	3790	83.3	918	22	AA151661 Triticum aestivum

10	2502.5	55.0	582	22	AA151670 Zea mays 101 kDa h
11	2431	53.4	668	22	AA151666 Arabidopsis thalia
12	2298	50.5	869	22	AA151667 Leishmania donovan
13	2289	50.3	867	22	AA151668 Leishmania major 1
14	2288	50.3	872	23	ABG77915 High level promote
15	2236.5	49.2	866	23	AB149337 Listeria monocytog
16	2224.5	48.9	868	22	AA151658 Trypanosoma brucei
17	2224	48.9	883	22	AA151651 Synchococcus CLPB
18	2204	48.4	861	22	AA151651 Salmonella typhi c
19	2194	48.2	857	20	AA151651 Staphylococcus aur
20	2193	48.2	857	22	AA151651 E. coli cellular p
21	2185	48.0	889	23	ABP66002 Bifidobacterium lo
22	2162.5	47.5	856	22	AA151651 Haemophilus influe
23	2158.5	47.4	848	23	ABJ04699 Mycobacterium tube
24	2145	47.1	852	22	AA151656 Corynebacterium gl
25	2145	47.1	852	22	AA151656 C glutamicum prote
26	2141.5	47.1	869	22	AA151656 Staphylococcus aur
27	2141.5	47.1	882	22	AA151656 Staphylococcus aur
28	2140.5	47.0	885	22	AA151656 Propionibacterium
29	2133	46.9	854	22	AA151656 Pseudomonas aerugi
30	2113	46.4	859	20	AA151656 Amino acid sequenc
31	2108	46.3	871	23	ABP39134 Staphylococcus epi
32	2107.5	46.3	859	24	ABP79777 N. gonorrhoeae ami
33	2099.5	46.1	867	23	AB154868 Lactococcus lactis
34	2033.5	44.7	977	22	AA151671 Phaseolus lunatus
35	1993.5	43.8	870	23	AA151671 Fungal heat resist
36	1938	42.6	867	22	AA151671 C. trachomatis CT1
37	1931	42.4	867	20	AA151671 Protein involved i
38	1912	42.0	866	22	AA151671 C. pneumoniae CT1
39	1912	42.0	872	20	AA151671 C. pneumoniae prot
40	1890	41.5	908	22	AA151657 Saccharomyces cere
41	1884.5	41.4	856	22	AA151657 Helicobacter pylor
42	1876.5	41.2	856	22	AA151657 Helicobacter pylor
43	1839	40.4	908	14	ABP33429 E. coli heat shock
44	1830	40.2	823	23	ABP39236 Staphylococcus epi
45	1828	40.2	817	22	AA151678 S. epidermidis ope

ALIGNMENTS

RESULT 1
AA151659
ID AA151659 standard; Protein; 911 AA.

AC AA151659;

XX

DT 14-JAN-2002 (first entry)

XX Arabidopsis thaliana heat shock protein 101 (GenBank: P42730).

DE Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX WO200170929-A2.

PD 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US08836.

XX 20-MAR-2000; 2000US-190769P.

XX 18-APR-2000; 2000US-198116P.

XX (ARCH-) ARCH DEV CORP.

XX Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.

XX N-PSDB; AA166065, AA166066.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA166057-AA166084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 911 AA;

Query Match 100.0%; Score 4550; DB 22; Length 911;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNPEKFTKNTTATTAHELAVNAGHAQTPHLAALISDPTGIFPOA1SSAGGENAAQ 60
 1 MNPEKFTKNTTATTAHELAVNAGHAQTPHLAALISDPTGIFPOA1SSAGGENAAQ 60
 61 SAERVINQALKLPSSPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLED 120
 61 SAERVINQALKLPSSPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLED 120
 121 SQRDLNNEVGATARVSEKELRGKGGKVESASGDTNFQALNTYGRDIVEQAGKLD 180
 121 SQRDLNNEVGATARVSEKELRGKGGKVESASGDTNFQALNTYGRDIVEQAGKLD 180
 181 VTGRDETRRVRIISRTKNNPVLIPGPGKTAIVGGLAQRIKVGDPVNSLTDVRLIS 240
 181 VTGRDETRRVRIISRTKNNPVLIPGPGKTAIVGGLAQRIKVGDPVNSLTDVRLIS 240
 241 LDMGALVAGAKYRGEFFERLKSVEDEDAEGKVLFTDEIHLVLGAGKTEGSDAANLF 300
 241 LDMGALVAGAKYRGEFFERLKSVEDEDAEGKVLFTDEIHLVLGAGKTEGSDAANLF 300
 301 KEMLARGOLRCIGATTLEBYRYKYVEKDAFERFOOVVVAEPPSDTTIILRGLEKYE 360
 301 KEMLARGOLRCIGATTLEBYRYKYVEKDAFERFOOVVVAEPPSDTTIILRGLEKYE 360
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVOLDSPQPEIDNLERK 420
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVOLDSPQPEIDNLERK 420
 421 MQLTELHALEREKDKASKARLIEYRKELDLDRDKLQPLTWKYRKEKERIDEIRLKKQR 480
 421 MQLTELHALEREKDKASKARLIEYRKELDLDRDKLQPLTWKYRKEKERIDEIRLKKQR 480
 481 BELMPSLOAERYDLARAADLRYCAI0EVESAIAOLEGTSSEENVMLTENVGPHIAEV 540
 481 BELMPSLOAERYDLARAADLRYCAI0EVESAIAOLEGTSSEENVMLTENVGPHIAEV 540
 541 VSRWTGIPVTRIGQNEKERLIGLADRLHKKRVVQGNQAVNAVSEALIRAGIGRAQOFTG 600
 541 VSRWTGIPVTRIGQNEKERLIGLADRLHKKRVVQGNQAVNAVSEALIRAGIGRAQOFTG 600
 601 SPLFLGPTGVGKTELAKALAEQFDENLLVRIDMSYMEQHSVRLIGAPPGYVGHGEG 660
 601 SPLFLGPTGVGKTELAKALAEQFDENLLVRIDMSYMEQHSVRLIGAPPGYVGHGEG 660
 661 GOLTEAVRRPYCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFRNSVIMTSN 720

Db 661 GOLTEAVRRPYCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFRNSVIMTSN 720
 QY 721 LGAHLLAGLTGKVTMEVARDCVMRVVRKHFRPELLNRLDEIVVDFPLSHDQLRKVARLQ 780
 Db 721 LGAHLLAGLTGKVTMEVARDCVMRVVRKHFRPELLNRLDEIVVDFPLSHDQLRKVARLQ 780
 QY 781 MKDVAVLAERGVVALAVTDAAIDYILAEISYDPVYGARPIRRWMEKKVTVLSKVVVEEII 840
 Db 781 MKDVAVLAERGVVALAVTDAAIDYILAEISYDPVYGARPIRRWMEKKVTVLSKVVVEEII 840
 QY 841 DENSTVIDAGDLVYRVESGLVDASTGKSDVLIHTANGPKRSDAAQAVKMMRIEII 900
 Db 841 DENSTVIDAGDLVYRVESGLVDASTGKSDVLIHTANGPKRSDAAQAVKMMRIEII 900
 QY 901 EDDNNEEMIED 911
 Db 901 EDDNNEEMIED 911

RESULT 2

AA166057
 ID AA166057 standard; Protein; 911 AA.

AA166084
 ID AA166084 standard; Protein; 911 AA.

14-JAN-2002 (first entry)

Triticum aestivum heat shock protein 101 (GenBank: AAD22629).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Triticum aestivum.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001MO-US08936.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

N-PSDB; AA166074.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA166057-AA166084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 911 AA;


```

Query Match      100.0%; Score 4550; DB 22; Length 911;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNPEKFTHTKTETATATAHELAVNAGHAQFTPLHLGALISDPTGIFFOAISSAGGENAAQ 60
1 MNPEKFTHTKTETATATAHELAVNAGHAQFTPLHLGALISDPTGIFFOAISSAGGENAAQ 60
61 SAERVINQALKKLPSSQSPDDIPASSSLIKVIRRAQAQKSRGDTNFAQKTYGRDLVQAGKLD 120
61 SAERVINQALKKLPSSQSPDDIPASSSLIKVIRRAQAQKSRGDTNFAQKTYGRDLVQAGKLD 120
121 SQIRDLLNEVGAVATARKVSEVKLRGKGGKVESASGDTNFAQKTYGRDLVQAGKLD 180
121 SQIRDLLNEVGAVATARKVSEVKLRGKGGKVESASGDTNFAQKTYGRDLVQAGKLD 180
181 VIGRDEEIRRVVRLSRRTKKNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
181 VIGRDEEIRRVVRLSRRTKKNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
241 LDMGALVAGAKYRGFEFERLKSVLKVEDEAGKVFILFIDEHLVLGAGKTEGSDMAANLF 300
241 LDMGALVAGAKYRGFEFERLKSVLKVEDEAGKVFILFIDEHLVLGAGKTEGSDMAANLF 300
301 KPMARGQLRCIGATTILEEYRKVYKVKDAAPRRFQQVYVAEPSVPTDISILGLKEKYE 360
301 KPMARGQLRCIGATTILEEYRKVYKVKDAAPRRFQQVYVAEPSVPTDISILGLKEKYE 360
361 HHGVRIQDRALINAQISARVITGRHLPDKAIDLVDEACANVRVQDLSQPEEDINLER 420
361 HHGVRIQDRALINAQISARVITGRHLPDKAIDLVDEACANVRVQDLSQPEEDINLER 420
421 MQLBIELHALEREKDKASKARLIEVRKELDDRLKQLPTMKYRKEKERIDEIRLKQKR 480
421 MQLBIELHALEREKDKASKARLIEVRKELDDRLKQLPTMKYRKEKERIDEIRLKQKR 480
481 EELMFSQAEARRVDLARAADLRVGAIOVESAIQAQEGTSSEENVMLTENVGPEHIAEV 540
481 EELMFSQAEARRVDLARAADLRVGAIOVESAIQAQEGTSSEENVMLTENVGPEHIAEV 540
541 VSRWTGIPVTRLGQNEKERLIGLADRLHKRVGQONAVNAVSAIRSRAGLGRAQOPTG 600
541 VSRWTGIPVTRLGQNEKERLIGLADRLHKRVGQONAVNAVSAIRSRAGLGRAQOPTG 600
601 SFLFLGPTGVGKTELAKALAEQLPDDENLLVRIDMSEYMEQHSVSRILGAPPGVVGHEG 660
601 SFLFLGPTGVGKTELAKALAEQLPDDENLLVRIDMSEYMEQHSVSRILGAPPGVVGHEG 660
661 GOLTEAVRRPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGGRVTPFRNSVITNSN 720
661 GOLTEAVRRPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGGRVTPFRNSVITNSN 720
721 LGAEHLLAGLTGKVTMEVARDQVMEVRKFRPELNLRLDEIVVEDPLSHDQLRKVARLQ 780
721 LGAEHLLAGLTGKVTMEVARDQVMEVRKFRPELNLRLDEIVVEDPLSHDQLRKVARLQ 780
781 MKDVAVRLAERGVAVLAVTDALDYVILAESYDPVYGARPIRRMWEKKVVTLSKQVREI 840
781 MKDVAVRLAERGVAVLAVTDALDYVILAESYDPVYGARPIRRMWEKKVVTLSKQVREI 840
841 DENSTVYIDAGAGDLVYRVESGGLVDASTGKSDVLIHIANGPKRSDAAQVKMKRIE 900
841 DENSTVYIDAGAGDLVYRVESGGLVDASTGKSDVLIHIANGPKRSDAAQVKMKRIE 900
901 EDDNEEMIED 911
901 EDDNEEMIED 911

ABP81189;
27-FEB-2003 (first entry)
Arabidopsis thaliana protein #17 modulated by PTGS.
Posttranscriptional gene silencing; PTGS; plant; transformation.
Arabidopsis thaliana.
W0200281695-A2.
17-OCT-2002.
05-APR-2002; 2002WO-EP03806.
06-APR-2001; 2001US-282049P.
(SYGN ) SYNGENTA PARTICIPATIONS AG.
(FRIE-) FRIEDRICH MIESCHER INST.
Zhu T, Glazov EA, Meins F, Wang X, Chang H;
WPI; 2003-103337/09.
N-PSDB; AB242033.
Novel polynucleic acid segment useful for modulating gene expression
within a cell by posttranscriptional gene silencing, and for augmenting
a plant cell genome -
Claim 53; Page 186-190; 438pp; English.
The invention relates to a novel isolated polynucleic acid segment
modulated within a cell by posttranscriptional gene silencing (PTGS). The
invention specifically relates to a method to identify an expression
product that is modulated by PTGS. The polynucleotide is useful for
modulating the gene expression within a cell by PTGS, by introducing the
polynucleic acid into a cell and expressing the nucleic acid segment in
the cell to form a product. The polynucleic acid segment is also useful
for augmenting a cell genome, and for augmenting a plant genome, by
contacting a plant cell with the segment to produce a transformed plant
cell, and growing the transformed plant cell to produce a differentiated
transformed plant. The sequences shown in ABP81173 - ABP81298 represent
the product of a segment of A. thaliana cDNA modulated by PTGS.
Sequence 911 AA;
Query Match      99.9%; Score 4545; DB 24; Length 911;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 910; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNPEKFTHTKTETATATAHELAVNAGHAQFTPLHLGALISDPTGIFFOAISSAGGENAAQ 60
Db 1 MNPEKFTHTKTETATATAHELAVNAGHAQFTPLHLGALISDPTGIFFOAISSAGGENAAQ 60
QY 61 SAERVINQALKKLPSSQSPDDIPASSSLIKVIRRAQAQKSRGDTNFAQKTYGRDLVQAGKLD 120
Db 61 SAERVINQALKKLPSSQSPDDIPASSSLIKVIRRAQAQKSRGDTNFAQKTYGRDLVQAGKLD 120
QY 121 SQIRDLLNEVGAVATARKVSEVKLRGKGGKVESASGDTNFAQKTYGRDLVQAGKLD 180
Db 121 SQIRDLLNEVGAVATARKVSEVKLRGKGGKVESASGDTNFAQKTYGRDLVQAGKLD 180
QY 181 VIGRDEEIRRVVRLSRRTKKNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
Db 181 VIGRDEEIRRVVRLSRRTKKNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
QY 241 LDMGALVAGAKYRGFEFERLKSVLKVEDEAGKVFILFIDEHLVLGAGKTEGSDMAANLF 300
Db 241 LDMGALVAGAKYRGFEFERLKSVLKVEDEAGKVFILFIDEHLVLGAGKTEGSDMAANLF 300
QY 301 KPMARGQLRCIGATTILEEYRKVYKVKDAAPRRFQQVYVAEPSVPTDISILGLKEKYE 360

```

301 KPMALGQRCIGATTLEEKYKVEKDAFERFQOYVVAEPSPDTISILRLKKEKYE 360
 361 HHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 420
 361 HHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 420
 421 MQLEIELHALEREKDKASARLIEVREKEDDLRLDKLOPLTMKYRKEKERIDIRLKKQR 480
 421 MQLEIELHALEREKDKASARLIEVREKEDDLRLDKLOPLTMKYRKEKERIDIRLKKQR 480
 481 EELMFSLQAEERYDLARAADRYGAIQVESALAEQLEGTSEENVMLTENVGPHIAEV 540
 481 EELMFSLQAEERYDLARAADRYGAIQVESALAEQLEGTSEENVMLTENVGPHIAEV 540
 541 VSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSRAGLGRQQPTG 600
 541 VSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSRAGLGRQQPTG 600
 601 SELFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSYMEQHSVSRLLICAGPGYVGHGEG 660
 601 SELFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSYMEQHSVSRLLICAGPGYVGHGEG 660
 661 GQTEAVRRRPPYCVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDNRNVLIMTSN 720
 661 GQTEAVRRRPPYCVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDNRNVLIMTSN 720
 721 LGAEHLLAGLTKVMTMEVARDCMREVRKHPPELLNRLDEIVDPDLSHQDLKRVARLQ 780
 721 LGAEHLLAGLTKVMTMEVARDCMREVRKHPPELLNRLDEIVDPDLSHQDLKRVARLQ 780
 781 MKDVAVRLAERGVAVLAVTDAALDYILAESYDVPYGARPIRRWMEKKVTELKMKVVRREI 840
 781 MKDVAVRLAERGVAVLAVTDAALDYILAESYDVPYGARPIRRWMEKKVTELKMKVVRREI 840
 841 DENSTVYIDAGGLVYRVESGGLVDASTGKSDVLIHANGPKHSDAAQVKKWRIE 900
 841 DENSTVYIDAGGLVYRVESGGLVDASTGKSDVLIHANGPKHSDAAQVKKWRIE 900
 901 EDDNEEMIED 911
 901 EDDNEEMIED 911

ULT 4
 AAM51662

AAM51662 standard; Protein; 911 AA.

AAM51662;

14-JAN-2002 (first entry)

Glycine max heat shock protein (GenBank: AAM51662).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Glycine max.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitech C, Vierling E;

DR WPI; 2001-639123/73.
 DR N-PSDB; AAI66067.
 PT Transgenic plants with improved heat stress tolerance, useful for
 PT producing animal feed, oil and synthetic products -
 XX Claim 2; Page -; 91pp; English.
 PS The invention relates to a transgenic plant, comprising a genetic
 XX construct comprising a promoter operatively linked to a nucleic acid
 CC sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 CC increased stress tolerance, especially to heat. The plant is a cereal,
 CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 CC synthetic product-producing plant, an environmental waste absorbing
 CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 CC an animal feed plant. In particular, the transgenic plant is cotton,
 CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 CC Arabidopsis thaliana. The plants may be used to produce animal feed,
 CC alcohol, crop, oil, medicine or a synthetic product.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.
 XX Sequence 911 AA;

Query Match 88.0%; Score 4003; DB 22; Length 911;
 Best Local Similarity 87.0%; Pred. No. 1.5e-289;
 Matches 794; Conservative 72; Mismatches 39; Indels 8; Gaps 6;

QY 1 MNEKFTHTKNTTATTAHNAVAGHAQFTPLHAGALISDPTGIFPQAISAA-GGENAA 59
 DB 1 MNEKFTHTKNTTATTAHNAVAGHAQFTPLHAGALISDPTGIFPQAISAA-GGENAA 59
 QY 60 QSAERVINQALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGTHLAVDOLIMLLE 119
 DB 61 RAVERVINQALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGTHLAVDOLIMLLE 119
 QY 120 DSQIRDLMEVGVATARVKSVEKLGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLD 179
 DB 121 DSQIGDLLEAGVAVAKVESEVDKLGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLD 180
 QY 180 PVIGRDEETRRVRIILSRRTKNNPVILGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 DB 181 PVIGRDEETRRVRIILSRRTKNNPVILGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLI 240
 QY 240 SLDMGALVAGAKYRGEPEERLKSVLKEVEDAGSKVILFIDEIHLVLGAGKTGSGMDAANL 299
 DB 241 ALDMGALVAGAKYRGEPEERLKSVLKEVEDAGSKVILFIDEIHLVLGAGKTGSGMDAANL 300
 QY 300 FKPMALARGQLRCIGATTLEEKYKVEKDAFERFQOYVVAEPSPDTISILRLKKEKYE 359
 DB 301 FKPMALARGQLRCIGATTLEEKYKVEKDAFERFQOYVVAEPSPDTISILRLKKEKYE 360
 QY 360 HHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 419
 DB 361 HHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 420
 QY 420 RMQLEIELHALEREKDKASARLIEVREKEDDLRLDKLOPLTMKYRKEKERIDIRLKKQR 479
 DB 421 RMQLEIELHALEREKDKASARLIEVREKEDDLRLDKLOPLTMKYRKEKERIDIRLKKQR 480
 QY 480 REELMFSLQAEERYDLARAADRYGAIQVESALAEQLEGTSEENVMLTENVGPHIAE 539
 DB 481 REELMFSLQAEERYDLARAADRYGAIQVESALAEQLEGTSEENVMLTENVGPHIAE 539
 QY 540 VSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSRAGLGRQQPT 599
 DB 540 VSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSRAGLGRQQPT 599
 QY 600 GSFLFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSYMEQHSVSRLLICAGPGYVGHGEE 659
 DB 600 GSFLFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSYMEQHSVSRLLICAGPGYVGHGEE 659

660 GQQLTEAVRRPVCVILFDEVEKAHVAVNTLLQVLDGRLTDGQRTVDFRNSVIMTS 719
 660 GQQLTEAVRRPVCVILFDEVEKAHVAVNTLLQVLDGRLTDGQRTVDFRNSVIMTS 719
 720 NLGASHLLAGLTGKVTMEVARDVMEVREKHFRELLNRLDEIVVFDPLSHDQRLKVARL 779
 720 NLGASHLLAGLTGKVTMEVARDVMEVREKHFRELLNRLDEIVVFDPLSHDQRLKVARL 779
 780 QMKQVAVRLAERGVAVLTDAAALDYLAESYDPVVGARPIRRWMEKKVTVTSLSKWVREE 839
 780 QMKQVAVRLAERGVAVLTDAAALDYLAESYDPVVGARPIRRWMEKKVTVTSLSKWVREE 839
 840 IDENSTVVIDAG--AGDLVYVE--SGGLVDASTGKSDVLIHIANG--PKSDAAQAVKK 894
 840 IDENSTVVIDAG--AGDLVYVE--SGGLVDASTGKSDVLIHIANG--PKSDAAQAVKK 894
 895 MRIEIEDDDNEE 907
 899 MKIEIEDDDNEE 911

ULT 5

AAM51665 standard; Protein; 909 AA.

AAM51665;

14-JAN-2002 (first entry)

Nicotiana tobacum 101 kDa heat shock protein (GenBank: AAC83688).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Nicotiana tobacum.

W0200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

N-PSDB; AAI66068.

Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AA166057-AA166084) encoding a plant Heat Shock protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number

CC reference provided in the specification.

XX	Sequence	909 AA;	Query Match	85.6%;	Score	3896;	DB	22;	Length	909;
XX	Best Local Similarity	85.0%;	Pred. No.	1.5e-281;						
XX	Matches	774;	Conservative	75;	Mismatches	56;	Indels	6;	Gaps	5;

QY	1	MNPEKTHKTHNETIATAHELVNAGHAQFTPLHLAGALLSDPTGIFPOAISSAGG-ENAA	59
DB	1	MNPEKTHKTHNETIATAHELVNAGHAQFTPLHLAGALLSDPTGIFPOAISSAGG-ENAA	60
QY	60	QSAERYINQAKKLPQSPPDPDIPASSSLIKVIRRAQAAQKSRGDTLHVLAVDQLMGLLE	119
DB	61	NSVERVLNQAKKLPQSPTAPDEIPSTSLIKVLRRAQSSQKSRGDTLHVLAVDQLMGLLE	120
QY	120	DSQIRDLNEVGVAHVAVKSEVEKRGKGVESASGDTNFOALKTYYGRDLVEQAGKLD	179
DB	121	DSQIGDLLKEAGVSASRVSEVEKRGKGVESASGDTNFOALKTYYGRDLVEQAGKLD	180
QY	180	PVIGRDEEIRRVVRIILSRRTKNNPVILGEPGVGKTAVVEGLAQRIVKGVDPNSLTDVRLI	239
DB	181	PVIGRDEEIRRVVRIILSRRTKNNPVILGEPGVGKTAVVEGLAQRIVKGVDPNSLTDVRLI	240
QY	240	SLDMGALVAGAKYRGFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTEGSDAANL	299
DB	241	ALDMGALVAGAKYRGFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTEGSDAANL	300
QY	300	FKPMLARGQLRCIGATTLEBYRYKVEKDAAPERRFQVYVAEPSVPTISILRGKKEYE	359
DB	301	FKPMLARGQLRCIGATTLEBYRYKVEKDAAPERRFQVYVAEPSVPTISILRGKKEYE	360
QY	360	GHGVKIQDRALINAAQLSARVITGRHLDPDKAIDLVDRACANRVQLDSQPEEDINLRRK	419
DB	361	GHGVKIQDRALINAAQLSARVITGRHLDPDKAIDLVDRACANRVQLDSQPEEDINLRRK	420
QY	420	RMQLELHALEREKDKASKARLIEVRKELDDRLDKLQPLTMKYKKEKERIDEIRRLKQK	479
DB	421	RIQLELHALEREKDKASKARLIEVRKELDDRLDKLQPLTMKYKKEKERIDEIRRLKQK	480
QY	480	REELMFPSIQEARRYDLARAADLRIGATQVEISAIQLEGTSSSEENVMLTENVGPEHIAE	539
DB	481	RDELTALQEAERRYDLARAADLRIGATQVEISAIQLEGTSSSEENVMLTENVGPEHIAE	539
QY	540	VYSRWITGIPVTRLGONEKERLIGLADRLHKRVGQNAVNAVSRALLESRAGLGRPOOPT	599
DB	540	VYSRWITGIPVTRLGONEKERLIGLADRLHKRVGQNAVNAVSRALLESRAGLGRPOOPT	599
QY	600	GSFLFLGPTGVGKTELAKALAEQOLFDDDKLMIRIDMSEYMEQHSVRLIGAPPGYVGHDE	659
DB	600	GSFLFLGPTGVGKTELAKALAEQOLFDDDKLMIRIDMSEYMEQHSVRLIGAPPGYVGHDE	659
QY	660	GGQLTEAVRRPVCVILFDEVEKAHVAVNTLLQVLDGRLTDGQRTVDFRNSVIMTS	719
DB	660	GGQLTEAVRRPVCVILFDEVEKAHVAVNTLLQVLDGRLTDGQRTVDFRNSVIMTS	719
QY	720	NLGAELHLAGLTGKVTMEVARDVMEVREKHFRELLNRLDEIVVFDPLSHDQRLKVARL	779
DB	720	NLGAELHLAGLTGKVTMEVARDVMEVREKHFRELLNRLDEIVVFDPLSHDQRLKVARL	779
QY	780	QMKQVAVRLAERGVAVLTDAAALDYLAESYDPVVGARPIRRWMEKKVTVTSLSKWVREE	839
DB	780	QMKQVAVRLAERGVAVLTDAAALDYLAESYDPVVGARPIRRWMEKKVTVTSLSKWVREE	839
QY	840	IDENSTVVIDAG--DLVYVE--SGGLVDASTGKSDVLIHIANGPKRSDAAQAVKKR	896
DB	840	IDENSTVVIDAG--DLVYVE--SGGLVDASTGKSDVLIHIANGPKRSDAAQAVKKR	896
QY	897	IEEIEDDDNEE 907	
DB	899	IEEIEDDDNEE 909	

RESULT 6

AAM51663

AAM51663 standard; Protein; 912 AA.

AAM51663;

14-JAN-2002 (first entry)

Zea mays heat shock protein 101 (GenBank: AAD25223).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Zea mays.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

N-PSDB; AA166072.

Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.

Sequence 912 AA;

Query Match 85.4%; Score 3883.5; DB 22; Length 912;

Best Local Similarity 84.6%; Pred. No. 1.3e-280;

Matches 772; Conservative 76; Mismatches 58; Indels 7; Gaps 6;

1 MNPEKTHKNTETATTAHELAVNAGHAQFTPLHLAGALISDPTGIFPOA1SSA-GGENAA 59

1 MNPDNTHKNTETATTAHELAVNAGHAQFTPLHLAGALISDPTGIFPOA1SSA-GGENAA 60

60 -QSAERVINQALKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTLHVDQLIMGLL 118

61 GDSFERNLNNLKLKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTLHVDQLIMGLL 120

119 EDSQIDLLNEGVATARYKSEVKLRGEGKKVSASGDTNFQALKTYGRDLVEQAKL 178

121 EDSQSDCLKEAGVSAARVRALEKLRGEGGRVESASGDTNFQALKTYGRDLVEQAKL 180

179 DPVIGRDEIRVRVILSRRTKNNPVLIGEPGVGKTA VVEGLAQRI VGDVPNSLTDVRL 238

Db 181 DPVIGRDEIRVRVILSRRTKNNPVLIGEPGVGKTA VVEGLAQRI VGDVPNSLTDVRL 240
 Qy 239 ISLDMGALVAGAKYRGFEERLKS VLVKEVEDAGKVLFDIHLVLGAKTEGSMDAAN 298
 Db 241 IALDMGALVAGAKYRGFEERLKA VLVKEVEEAGKVLFDIHLVLGAKTEGSMDAAN 300
 Qy 299 LFKPMLARGQLRCIGATTILBEYRK YVEKDAAFERRFOQVVAEFPVDTISILRGLKEKY 358
 Db 301 LFKPMLARGQLRCIGATTILBEYRK YVEKDAAFERRFOQVVAEFPVDTISILRGLKEKY 360
 Qy 359 EGHGVRIQDRALINAAQLSARYITG RHLDPKADLVDEACANRVQLDSQPEIDNLER 418
 Db 361 EGHGVRIQDRALVVAQAQLSARYITG RHLDPKADLVDEACANRVQLDSQPEIDNLER 420
 Qy 419 KRMOLTEHLALREKDKASKARLIEVR KELDLDLKDLOPLTMKYRKEKERIDEIRLQK 478
 Db 421 KRIQIEVELHALEKEKDKASKARLIEVR KELDLDLKDLOPLTMKYRKEKERIDEIRLQK 480
 Qy 479 KREELMFSLQEAERRYDLARAADIRYGA IOEVSASIAQLSEGTSEENVMLTENVGPEHIA 538
 Db 481 RREELQFTLQEAERRMDLARVADLKYGA LQEIIDAALSKLE-SETGENMLTETVGEQIA 539
 Qy 539 EVSRWTGIPVTRLGQNEKERLIGLADRL HKRVTGQNAVNASEAILRSRAGIGRAQOP 598
 Db 540 EVSRWTGIPVTRLGQNDKERLVGLADRL HQRVVGQTEAVSAVAEAVLRSRAGIGRPOOP 599
 Qy 599 TGSFLFLGPTGVGKTELAKALAEQLPDD ENLLVRIDMSEYMEQHSVSRLLIGAPGVGHE 658
 Db 600 TGSFLFLGPTGVGKTELAKALAEQLPDD ENLLVRIDMSEYMEQHSVSRLLIGAPGVGHE 659
 Qy 659 EGGQLTEAVRRRYPYCVILFDEVEKAHVA VENTLLQVLDDGRLTDGQRTVDFRNSVIIMT 718
 Db 660 EGGQLTEQVRRRYPYCVILFDEVEKAHVA VENTLLQVLDDGRLTDGQRTVDFRNTVIIMT 719
 Qy 719 SNLGAELHLAAGLTGKVTMEVARDCVWR EKRFRPELLNRLDEIVVDFPLSHDQLKVAR 778
 Db 720 SNLGAELHLAAGLVGKSNMKVARDLVNQ EVRHRFRPELLNRLDEIVVDFPLSHDQLKVAR 779
 Qy 779 LQMKDVAVRLAERGVAVLAVTDAALDYLA ESYDPVYVGARPIRRWMEKKVVTLSKMWVRE 838
 Db 780 LQMKDVAVRLAERGVAVLAVTDAALDYLA ESYDPVYVGARPIRRWMEKKVVTLSKMWVRE 839
 Qy 839 EIDENSTVYIDAGAG--DLVYRVE--SGGL VDASTGKSDVLHIANGPKRSDAQAQVKM 895
 Db 840 EIDENCTVYIDAPGKDELVYRDRSGGLVNA ETKMSDILLIQVPNSSTRSDAAQAQVKM 899
 Qy 896 RI-REIEDDDNEE 907
 Db 900 RIMEEDEDGMDEE 912

RESULT 7
 AAM51669
 ID AAM51669 standard; Protein; 912 AA.
 XX AAM51669;
 AC AAM51669;
 XX
 DT 14-JAN-2002 (first entry)
 XX
 DE Zea mays heat shock protein HSP101 (GenBank: AAD33606).
 XX
 KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana.
 XX
 OS Zea mays.
 XX
 FN WO200170929-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 20-MAR-2001; 2001WO-US08836.
 XX

10- KXETZQ 1207REINQNDENK VADUURIGATQEDDAIISNLE-SEIGENLMLTETVGPQIA 539

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA160597-AA16084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA551651-AA551671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing

plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 913 AA;

Query Match 84.0%; Score 3822.5; DB 22; Length 913;
Best Local Similarity 83.3%; Pred. No. 4.5e-276;
Matches 761; Conservative 75; Mismatches 69; Indels 9; Gaps 6;

```

1  MNPEKTHKTNETIATAHELAVNAGHAQTPHLHAGALISDPTGIFPOAISAGGEN--A 58
1  MNPKDTHKTNELAAHEWASEAGHAQTUPLHAAALADRSGLRQALHAGSGNDAA 60
59  AQAERVINOALKKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL 118
61  AESFERVAGALKRPLSPQSPDDTVPASTSLVKAVERAQAQKSRGDSHLAVDQIMGLL 120
119  EDSQIDILNEGVATARKVSEVKRGKGGKVVESASGDTNFOALKTYGRDLVEVAGKL 178
121  EDQISDALKEAGISAARKAEYKELRGDNRVESASGDTNFOALKTYGRDLVEVAGKL 180
179  DPVIGDEIRRVVRLSRRTKONPVLLGPGVGTAVVEGLAQRVKGVDVNSITDVLRL 238
181  DPVIGDEIRRVVRLSRRTKONPVLLGPGVGTAVVEGLAQRVKGVDVNSITDVLRL 240
239  ISLDMGALVAGKRGFEFERLKSILKVEDAEGKVLFDIHLVLAGKTEGSDMAAN 298
241  VALDMGALVAGKRGFEFERLKSILKVEDAEGKVLFDIHLVLAGKTEGSDMAAN 300
299  LKPMALARGQLCIGATTLEERKVKYKDKAAFERRFQVYVAEPSVPTISLRLGKRY 358
301  LKPMALARGQLCIGATTLEERKVKYKDKAAFERRFQVYVAEPSVPTISLRLGKRY 360
359  EGHGVRIQDRALINAAQLSARYITGRHLDPKADLVDACANVRVOLDSPEDINLER 418
361  EGHGVRIQDRALINAAQLSARYITGRHLDPKADLVDACANVRVOLDSPEDINLER 420
419  KMOLEIEHLERKOKKASKARLIEVRKELDDLKPLTMKYRKEKERIDEIRLKQ 478
421  KRIQLEVLHAEKEDKASKARLVDRKELDDLKPLTMKYRKEKERIDEIRLKQ 480
479  KEELMFLSQEAEERYDLARAADLRYGATQEVESATQLEGTSSEENVMTENVGPEHIA 538
481  REELQFTLOEAEERRMDLARVADLRYGALQEVDAALKEGTG-ENLMLETVTGPDQIA 539
539  EVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVGQNOAVNAVSEAILRSRAGLRGQOP 598
540  EVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVGQNOAVNAVSEAILRSRAGLRGQOP 599
599  TGSFLFLGTGKTELAKALAEQIPDENLLVIRIDMEYMEQHSVSLIGAPPYVGH 658
600  TGSFLFLGTGKTELAKALAEQIPDENLLVIRIDMEYMEQHSVSLIGAPPYVGH 659
659  EGQQLTEAVRRPYPYVILLFDEVEKHAFAVNTLLQVLDGRLTDCQGRVDFRNSVIIMT 718
660  EGQQLTEAVRRPYPYVILLFDEVEKHAFAVNTLLQVLDGRLTDCQGRVDFRNSVIIMT 719
719  SNLGAHLLAGTGTGKTEVARDVCMREVRKHFRPELNLRLDEIVVFDPLSHDQLRKVAR 778
720  SNLGAHLLAGTGTGKTEVARDVCMREVRKHFRPELNLRLDEIVVFDPLSHDQLRKVAR 778
779  LOMKDVAVRLAERGVALAVTDALDIYILAESVDVPYVGARPIRRMWEKVVTELKVVRE 838
779  LOMKDVAVRLAERGVALAVTDALDIYILAESVDVPYVGARPIRRMWEKVVTELKVVRE 838
839  EIDENSTVYIDAGAG--DLVYRVES--GGLVDASTGKSDVLHLIANGPKSRDAQAQVKM 895

```

Db 839 EIDENSTVYIDAAERSKDELTVGVKHGELVNARTGHKSDILIQVPSGAVGGDAHAHVKKM 898

QY 896 RI--BEIEDDDNEE 907

Db 899 KIMQDSGEVDDEE 912

RESULT 9

AAM51661

XX AAM51661 standard; Protein; 918 AA.

XX AC AAM51661;

XX 14-JAN-2002 (first entry)

XX Triticum aestivum 101kDa heat shock protein (GenBank: AAC83689).

XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX Arabidopsis thaliana.

XX Triticum aestivum.

XX WO200170929-A2.

XX 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US08836.

XX 20-MAR-2000; 2000US-190769P.

XX 18-APR-2000; 2000US-198116P.

XX (ARCH-) ARCH DEV CORP.

XX Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.

XX N-PSDB; AAI66075.

XX Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -

XX Claim 2; Page -; 91pp; English.

XX The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal,
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.

XX Note: The sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

XX Sequence 918 AA;

Query Match 83.3%; Score 3790; DB 22; Length 918;

Best Local Similarity 81.7%; Pred. No. 1.2e-273;

Matches 752; Conservative 86; Mismatches 70; Indels 12; Gaps 7;

QY 1 MNPEKTHKTNETIATAHELAVNAGHAQTPHLHAGALISDPTGIFPOAISAGGENAA- 59

Db 1 MNPDNFTKTNELAAHEWASEAGHAQTUPLHAAALADRSGLRQALHAGSGNDAA 60

QY 60 -QSAERVINOALKKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL 118

Db 61 GDSFERVLGALRLKPLSPQSPDDTVPASTSLVKAVERAQAQKSRGDSHLAVDQIMGLL 120

119 EDSQIDLLNEGVATARKVEKLGKE-GKKVESASGDTNFQALKTYGRDLVEQAGK 177
 121 EDAQADCLKEAGVSASRVARELKLGGDSNRKVSFESAGDFTFOALKTYGRDLVEVAGK 180
 178 LDPVIGRDEIRRVRIILSRRTKNNPVLIGEPGVGTAVVEGLAORIVKGVDPNSLTDVR 237
 181 LDPVIGRDEIRRVRIILSRRTKNNPVLIGEPGVGTAVVEGLAORIVKGVDPNSLTDVR 240
 238 LISLDMGALVAGAKYRGEFEELKSVLKEVEDAEGKVLFIIDEIHLVLGAGKTEGMDAA 297
 241 LVALDMGALVAGAKYRGEFEELKSVLKEVEDAEGKVLFIIDEIHLVLGAGKTEGMDAA 300
 298 NLFKPMLAGQLRCGATTLEBYRKYRVEKDAAFERRFOQVYVAEPPDTSILRLGKKEK 357
 301 NLFKPMLAGQLRCGATTLEBYRKYRVEKDAAFERRFOQVYVAEPPDTSILRLGKKEK 360
 358 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLE 417
 361 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLE 420
 418 RKRMOLETELHALEKEKASKARLIEVRKELDDLDLRLKQLPTMKYRKEKERIDEIRLK 477
 421 RKRMOLETELHALEKEKASKARLIEVRKELDDLDLRLKQLPTMKYRKEKERIDEIRLK 480
 478 OKREELMFSLOEABERYDLARAADLYCAI QEVESAIQAQLEGTSEENVMLTENVPBEHI 537
 481 OKREELMFSLOEABERYDLARAADLYCAI QEVESAIQAQLEGTSEENVMLTENVPBEHI 539
 538 AEVSRWTGIPVTRLGQNKERLIGLADRLHKKRVVQGNQAVNAVSEATLRAGLGRQQ 597
 540 AEVSRWTGIPVTRLGQNKERLIGLADRLHKKRVVQGNQAVNAVSEATLRAGLGRQQ 599
 598 PTGSFLPGPTGVGKTELAKALAEQFDENLVRIDMSYMEQHSVRLTGAPPGYVGH 657
 600 PTGSFLPGPTGVGKTELAKALAEQFDENLVRIDMSYMEQHSVRLTGAPPGYVGH 659
 658 EGGQLETAVERRRPVCVILEVEKAVHVAVENTLLOVLDGRLTDGQRTVDFRNSVIM 717
 660 EGGQLETAVERRRPVCVILEVEKAVHVAVENTLLOVLDGRLTDGQRTVDFRNSVIM 719
 718 TSNLGAELLAGLTKVMEVARDVMEVREKRPPELLNRLDEIVVDFDLSHDLKRA 777
 720 TSNLGAELLAGLTKVMEVARDVMEVREKRPPELLNRLDEIVVDFDLSHDLKRA 779
 778 RLOMKDVAVRLAERGVAVLAVTDAALDYILABSYPVYGARPIRMEKVKVTELKRMVVR 837
 780 RLOMKDVAVRLAERGVAVLAVTDAALDYILABSYPVYGARPIRMEKVKVTELKRMVVR 839
 838 EEIDENSTVYIDAGAGD-LVYRVE-SGLVDASTGKKSDVLIHTANGP-----KESDAAQA 891
 840 EEIDENSTVYIDAGAGD-LVYRVE-SGLVDASTGKKSDVLIHTANGP-----KESDAAQA 899
 892 VKKRIIEIEDDDNEEMTED 911
 900 VKKRIIEIEDDDNEEMTED 917

ULT 10
 AAM51670

AAM51670 standard; Protein; 582 AA.

AAM51670;

14-JAN-2002 (first entry)

Zea mays 101 kDa heat shock protein (GenBank: AAD26530).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Zea mays.

XX WO200170929-A2.
 XX 27-SEP-2001.
 XX 20-MAR-2001; 2001WO-US08836.
 XX 20-MAR-2000; 2000US-190769P.
 PR 18-APR-2000; 2000US-198116P.
 XX (ARCH-) ARCH DEV CORP.
 XX Lindquist S, Queitsch C, Vierling E;
 PI WPI; 2001-639123/73.
 DR N-PSDB; AAI66071.
 XX Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -
 Claim 2; Page -; 91pp; English.
 The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAMS1651-AAMS1671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.

Sequence 582 AA;

Query Match 55.0%; Score 2502.5; DB 22; Length 582;
 Best Local Similarity 85.2%; Pred. No. 7e-178;
 Matches 497; Conservative 49; Mismatches 32; Indels 5; Gaps 4;

QY 329 AFERRQQVVAEPSVPDTSILRLGKKEKVEGHGVRIQDRALINAAQLSARYITGRHLP 388
 Db 1 AFERRQQVVAEPSVPDTSILRLGKKEKVEGHGVRIQDRALINAAQLSARYITGRHLP 60
 QY 389 DKALDLDVDEACANVRVQLDSQPEEIDNLERKQMLEIEIHLALEREKDKASKARLIEVRKE 448
 Db 61 DKALDLDVDEACANVRVQLDSQPEEIDNLERKQMLEIEIHLALEREKDKASKARLIEVRKE 120
 QY 449 LDDLRLKQLPTMKYRKEKERIDEIRLKKQKEELMFSLOEABERYDLARAADLYCAI 508
 Db 121 LDDLRLKQLPTMKYRKEKERIDEIRLKKQKEELMFSLOEABERYDLARAADLYCAI 180
 QY 509 EVESAIQAQLEGTSEENVMLTENVPGEHIAEVVSWTGTIPVTRLGQNKERLIGLADRLH 568
 Db 181 EIDAAISKLE-SETGENMLTETVGEQIAEVVSWTGTIPVTRLGQNKERLIGLADRLH 239
 QY 569 KRVVQGNQAVNAVSEATLRAGLGRQQPTGSFLPGPTGVGKTELAKALAEQFDEN 628
 Db 240 QRVVQGTSAVSAVAEVLRLSRAGLGRQQPTGSFLPGPTGVGKTELAKALAEQFDEN 299
 QY 629 LLVRIDMSYMEQHSVRLIGAPPYGVGHEGGQLETAVERRRPVCVILEVEKAVHVA 688
 Db 300 LLVRIDMSYMEQHSVRLIGAPPYGVGHEGGQLETAVERRRPVCVILEVEKAVHVA 359
 QY 689 NTLLQVLDGRLTDGQRTVDFRNSVITSNLGAELLAGLTKVMEVARDVMEVREVR 748
 Db 360 NTLLQVLDGRLTDGQRTVDFRNSVITSNLGAELLAGLTKVMEVARDVMEVREVR 419
 QY 749 KHFRPELLNRLDEIVVDFDLSHDLKRAVRLAERGVAVLAVTDAALDYILAE 808

420 RHRPELNLDELDEIVIPDPLSHQELRKVARLQMKDVAVRLAERGIALAVTDAALDIISL 479
 809 SYDPVYGARPIRWMEKKVVTLSKVVREIENSTVYIDAGAG--DLVYRVE--SGGLV 865
 480 SYDPVYGARPIRWKIEKRVVTLQSKMLDIEIDENCTVYIDAAAGKDELIVYRVDSSGGLV 539
 866 DASTGKSDVLIHANGPKRSDAAQAVKQRI--EEIEDDDNEE 907
 540 NAETGMSDILIQVPTSTSRSDAAQAVKQRIEEDDEEDGMD 582

RESULT 11

AAM51666
 AAM51666 standard; Protein; 668 AA.

AAM51666;

14-JAN-2002 (first entry)

Arabidopsis thaliana heat shock protein like (GenBank: CAB46061).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Arabidopsis thaliana.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.
 18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.

Sequence 668 AA;

Query Match 53.4%; Score 2431; DB 22; Length 668;
 est Local Similarity 70.8%; Pred. No. 1.9e-172;
 atches 499; Conservative 76; Mismatches 90; Indels 40; Gaps 6;

1 MNPEKTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSAG-GENAA 59
 1 MNDLKFDPNVKLILASARSHAMLSHGQVTPHLGLVTLISLTSVFFYRAITSAGDGISA 60

QY 60 QSAERVINOALKKLPSQSPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQLIMGLLE 119
 DB 61 QSVNVIINQSLYKL-----TKRNLGDTKVGVAVLVLSLE 95
 QY 120 DSQIRDLILNEGVGATARVKSEVKLRGKKGKVBESAGDTNFQALKTYGRDLVEQAQKLD 179
 DB 96 DSQISDLVLEAGVVPKVKSEVKLR-----GEVILRAKTYGTDLVEQAQKLD 144
 QY 180 PVIGRDEIRVRVRLISRRTKQNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 DB 145 PVIGRHRIRRVIEVLSRRTKQNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNLITGVKLI 204
 QY 240 SLDMGALVAGAKYGEFEERLKSVLKEVEDAEGKVILFIDEIHLVLGAGTSGMSMDANL 299
 DB 205 SLEFGAMVAGTTLRGQFEERLKSVLKAVEBAQGVILFIDEIHMALGACKASGSDAAKL 264
 QY 300 FKPMIARGQLRCIGATTLEBYRVKYVEKDAAFERRFOQVVAEVPSPDTISILGLKEKYE 359
 DB 265 LKPMIARGQLRFIGATTLEBYRVTHVEKDAAFERRFOQVVAEVPSPDTISILGLKEKYE 324
 QY 360 GHGVRIQDRALINAAQLSARYITGRHLDPKALDLVDEACANVRVQLDSQPEEIDNLERK 419
 DB 325 GHGVRIQDRALVLSAQLSERVITGRRLPKALDLVDESAHVKAQLDIQPEEIDSLEK 384
 QY 420 RMQLEIELHALEREK-DKASKARLIEVRKELDDLRLDKLOPLTMKYRKEKERIDEIRLQK 478
 DB 385 VMQLEIEIHALEREKDKSEARLSEVRKELDDLRLDKLEPLTIKYKKEKKIINETRLQK 444
 QY 479 KRELMFSLQEAERRVYDLARAADLYGAIQEVESALAQLEGTSEENVMLTENVGPEHIA 538
 DB 445 NRDDLMTALQEAERQHDVPKAAVLYKGAIQEVESALAKLE-KSAKDNVMLTEIVGGENIA 503
 QY 539 EVYSRWTCIPIVTRLGQNEKERLIGLADRLHKRVVGQNAVNAVSEAILRSRAGLRAQOP 598
 DB 504 EVYSRWTCIPIVTRLDQNEKKRLISLADKLHERVVGQDEAVKAVAAAILRSRVLGRPOOP 563
 QY 599 TGSFLFLGPTGVGKTELAKALAEQLFDENLIVRIDMSEYMEQHSVSRLLIGAPPG-VYGH 657
 DB 564 SGSFLFLGPTGVGKTELAKALAEQLFDENLIVRLDMSEYNDKFSVNLKLGAPPGYIGH 623
 QY 658 EGGQLTEAVRRRPPYCVILFDEVEKARHVAVFNTLLQVLDGRLTD 702
 DB 624 EGGQLTEPVRRRPPYCVILFDEVEKTHVTVFNTLLQVLDGRLTD 668

RESULT 12

AAM51667

ID AAM51667 standard; Protein; 869 AA.

XX AC AAM51667;

XX DT 14-JAN-2002 (first entry)

XX DE Leishmania donovani heat shock protein 100 (GenBank: CAB08073).

XX KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana.

XX OS Leishmania donovani.

XX PN WO200170929-A2.

XX PD 27-SEP-2001.

XX PF 20-MAR-2001; 2001WO-US08836.

XX XX 20-MAR-2000; 2000US-190769P.

PR 18-APR-2000; 2000US-198116P.

XX PA (ARCH-) ARCH DEV CORP.

XX

Lindquist S, Queitsch C, Vierling B;

WPI; 2001-639123/73.

N-PSDB; AAI66083.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or plant, an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 869 AA;

Query Match 50.5%; Score 2298; DB 22; Length 869;
 Best Local Similarity 54.4%; Pred. No. 2.3e-162;
 Matches 462; Conservative 153; Mismatches 224; Indels 10; Gaps 5;
 3 PEKTHKNTETIAHVLAVNAGHAQTPPLHLAGALISDPTGIPPOAISSAGGENAQA 62
 6 PE-WTQAASDLMAKARKKANGYLDVHLAVVPEDENSLASRVKLG----AASV 60
 63 ERVINQALKLPSPSPDDIPASSSLIKVIRRAQAQKRGDTHLAVDQIMGLIEDSQ 122
 61 KQLEARVDAIPTQMPAPTQPRNSDMWRVMTAEQERVALGDTLMAADHFLALHESKE 120
 123 IRDLNVEGVVATARKVEKLRGKGEKGVESAGDTNFOALKTYGRDLVEQA--GKLD 180
 121 VGRILDAAGAKKAIATILEMR--KGGKINSFQDDNYESLNKYAVDLCKQAEKGKLD 178
 181 VIGRDEIRVRLSRRTKNNPVLIGEPGVGTAVVGLAQRIVKGDVPSNLTVDRLIS 240
 179 VIGRADILRTIRVLSRTKNNPVLIGEPGVGTAVVGLAQQVREGDPTLSGRIFS 238
 241 LDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDHILVLAGKTEGSDAANLF 300
 239 LDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDHILVLAGKSDGANDAANLL 298
 301 KPMIARGQLRCIGATTILEYRKVYKVDAAFERFQVYVAERPSVPTISILRGLKEKYG 360
 299 KPIILARGELATIGATTILEYRKVYKVDAAFERFQVYVAERPSVPTISILRGLKDRYEQ 358
 361 HHGVRQDRALINAQLSARYITGRHLPDKAIDLVDCAENRVQDSQPEEDINLERK 420
 359 HHGVTQDKAVVVAAGLAGRYITGRHLPDKAIDLVDCAENRVVLTSSPAEDALERK 418
 421 MQLEIEHLALEREKDKASKARLLEVRKELDDLKDLKLOPLTMKYRKEKREIDEIRLKQR 480
 419 RQLEIEKALQDKDSAKERLKAVERKAEIKQVEKGLPLLAQYQERGRIDELQATQAKL 478
 481 EELMFSIQEAERYDILARAADIRYGAIQEVESAIAOL--EGTSSEENVMLTENVEPHAE 539
 479 DEKKVKLAEARMWDMETADLKNVPILODRIRSLKEEIKQKATMLHGTVTETDIAT 538
 540 VSRWTGIPVTRIGQNEKERLIGLADRLHKKRVVQGNQAVNSEAILRSRAGIGRAQOFT 599
 539 VSRWTNIPVTKLSQTERELLHLDQLHLRVKQDEAVSRVAEAILRSRAGLSDRPT 598
 600 GSFLFGTGVGKTELAKALAEQLFDENLLVRIDMSEYMHQHSVSRSLTGAPPGYVGHHE 659

Db 599 GSFLFGTGVGKTELAKALAEQLFDENLLVRIDMSEYMHQHSVSRSLTGAPPGYVGHHE 658
 Qy 660 GGQLEAVRRPYCVILFDEVERKAHVAVFNTLLQVLDGRLTDGQGRITVDFRNSVIMTS 719
 Db 659 GGQLEAVRRPYCVILFDEVERKAHVAVFNTLLQVLDGRLTDGQGRITVDFRNSVIMTS 718
 Qy 720 NLGAHELLAGLTKGVTMEYVARDCYWEVRKHFPELNLRLDELIVFDPPLSHDQLRKVARL 779
 Db 719 NLGAHELLAGLTKGVTMEYVARDCYWEVRKHFPELNLRLDELIVFDPPLSHDQLRKVARL 778
 Qy 780 QMKDVAVRLAERGVAVLAVTDAALDYTLAESYDPVYVGARPLRRWMEKKVVTLSKMTVREE 839
 Db 779 ITEELVGLKQDSIRVSLTEERAKYVLESFADMGARFRRWVEKNITTELSRMIISOE 838
 Qy 840 IDENSTVYI 848
 Db 839 LSPNSTVKV 847
 RESULT 13
 AAM51668
 ID AAM51668 standard; Protein; 867 AA.
 XX AC AAM51668;
 XX 14-JAN-2002 (first entry)
 XX Leishmania major 100 kDa heat shock protein (GenBank: CAA86116).
 DE Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana.
 XX Leishmania major.
 OS WO200170923-A2.
 XX 27-SEP-2001.
 XX 20-MAR-2001; 2001WO-US08836.
 XX 20-MAR-2000; 2000US-190769P.
 PR 18-APR-2000; 2000US-198116P.
 XX (ARCH-) ARCH DEV CORP.
 PA Lindquist S, Queitsch C, Vierling B;
 PI WPI; 2001-639123/73.
 DR N-PSDB; AAI66084.
 XX Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -
 Claim 2; Page -; 91pp; English.
 The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or plant, an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 867 AA;

Query Match 50.3%; Score 2289; DB 22; Length 867;
 Best Local Similarity 54.3%; Pred. No. 1.1e-161;
 Matches 461; Conservative 151; Mismatches 229; Indels 10; Gaps 5;
 3 PEKFTHTKNTTATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAQAOS 62
 6 PE-WKQAASDLMAAALARKKANGVLDPLHVLAVYMFEDENSILASRAVRKLG---AASV 60
 63 ERVINQALKKLSQSPPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLS 122
 61 KQGLARVAIDINQMPATQPPNSDMFVMTAEQERAAALGDTLMAADHFLALHESKE 120
 123 IRDLNVEGVATARVSEVEKLRGKGVKVESASGDTNFQALKTYGRDLVEQA--GKLD 180
 121 VGRILDAAGAGKKAIRATILEWR--KGKITSDFQDNYESLNKYAVDLCKQAEKGKLD 178
 181 VIGRDEIRRVRLSRRTKNNPVLIGEGVGKTAVVEGLAQRIIVKGDVPSNLTDLVRLS 240
 179 VIGRADEILRTIRVLSRRTKNNPVLIGEGVGKTAIVEGIAQVVRGDPVDTLSGIRIFS 238
 241 LDMGALVAGAKYRGEFEERLKSILVEDEDAEGKVLFTDEIHLVLAGKTEGMDAANLF 300
 239 LDMGALVAGAKYRGEFEERLKSILVEDEDAEGKVLFTDEIHLVLAGKTEGMDAANLF 298
 301 KPMARGQLRCIGATTLEBYRYKYVEKDAFERFQVVAEPVDTTISILGLKKEG 360
 299 KPLLAGDVRTIGATTLEBYRYKYVEKDAFERFQVVAEPVDTTISILGLKKEG 358
 361 HGVRIQDRLINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPBEINDLERK 420
 359 HGVRIQDRLINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPBEINDLERK 418
 421 MOLETELHAREKQKASARLIEVREKELDLRDLKPLTKYRKEKERIDEIRLKKR 480
 419 ROLETELHAREKQKASARLIEVREKELDLRDLKPLTKYRKEKERIDEIRLKKR 478
 481 BELMFSLQAEARRYDLARAADLRGAIQVESAIAQL-EGTSSEENVMLTENVGPEHAE 539
 479 DEKKYLERAEARMDETAADLRGAIQVESAIAQL-EGTSSEENVMLTENVGPEHAE 538
 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHKKRVGQNAVSEAILRSRAGLGRAOQPT 599
 539 VVSRWTGIPVTRLGQNEKERLIGLADRLHKKRVGQNAVSEAILRSRAGLGRAOQPT 598
 600 GSFELGPTGVGKTELAKALABOLDDENLVRIDMSYMEQHSVRLIGAPGVVGHHE 659
 599 GSFELGPTGVGKTELAKALABOLDDENLVRIDMSYMEQHSVRLIGAPGVVGHHE 658
 660 GQQLTEAVRRPVCYVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDFRNSVIMTS 719
 659 GQQLTEAVRRPVCYVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDFRNSVIMTS 718
 720 NLGAHLLAGTGTMTVEVARDVCVMEVREKHFRELPBELLNDEIVFEDPLSHDQKRVAR 779
 719 NLGAHLLAGTGTMTVEVARDVCVMEVREKHFRELPBELLNDEIVFEDPLSHDQKRVAR 778
 780 QMKDQVRLAEAGVALVTDALDYLAEVDPVVGARPIRMEKVVITSLKVVREE 839
 779 ITEELNGLKQOSIRVSLTEBAKQVLESADFADMGARPLRRRMEKVVITSLKVVREE 838
 840 IDENSTVYI 848
 839 LSPNSTVKV 847

UL14
 ABG77915
 standard; Protein; 872 AA.
 ABG77915;
 ABG77915;

05-NOV-2002 (first entry)
 High level promoter polypeptide #19.
 High level promoter; Cyanobacterium; UV-B light; green alga; plant;
 sunlight biomaterial.
 Synechocystis sp. PCC6803.
 WO200261098-A2.
 08-AUG-2002.
 30-JAN-2002; 2002WO-US03926.
 30-JAN-2001; 2001US-264925P.
 (DUPO) DU PONT DE NEMOURS & CO E I.
 Huang LL, Larossa RA, McCluskey MP;
 WPI; 2002-619256/66.
 N-ESDB; ABS63272.
 Regulating gene expression of a coding region in Cyanobacterium, useful
 in identifying highly expressed or UV responsive genes and their
 promoters, comprises culturing the transformed cell in log phase or in
 the presence of UV-B light -
 Claim 4; Page 83-85; 86pp; English.
 The invention relates to a method for regulating gene expression of a
 coding region in a Cyanobacterium comprising culturing the transformed
 Cyanobacterium in the log phase or in the presence of UV-B light, where
 the promoter region is activated and the coding region is expressed. The
 method is useful for identification of highly expressed genes or UV
 responsive genes, and their corresponding promoters. These genes and
 promoters are useful for constructing expression vectors in
 Cyanobacteria, green algae or plants and for the production of
 biomaterials from sunlight. This sequence represents a high level
 promoter polypeptide of the invention.
 Sequence 872 AA;
 Query Match 50.3%; Score 2288; DB 23; Length 872;
 Best Local Similarity 52.6%; Pred. No. 1.3e-161;
 Matches 460; Conservative 167; Mismatches 229; Indels 16; Gaps 7;
 2 NPEKFTHTKNTTATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAQAOS 61
 5 DPNKFTKAWALAKTPEIAKQHRQOQIETHELLSALL-EQGLATSIFNKAGA-----S 58
 62 AERV---INQALKLSQSPPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGL 118
 59 IPRVNDQVNSFIAQPKLSNPSESIYLGSLDKLDNAEIAKSKYGGDDVISTEHLMAAYG 118
 119 EDSQI-RDILNVEGVATARVSEVEKLRGKGVKVESASGDTNFQALKTYGRDLVEQA-- 175
 119 QDRILGNLYREIGLTENKLABIIKQIRGTQ--KYTDQNPGEKYESLEKYGRDLTARE 176
 176 GKLDPVIGRDEEIRRVRLSRRTKNNPVLIGEGVGKTAIVVEGLAQRIIVKGDVPSNLT 235
 177 GKLDPVIGRDEEIRRVRLSRRTKNNPVLIGEGVGKTAIVVEGLAQRIIVKGDVPSNLT 236
 236 VRLISLDMGALVAGAKYRGEFEERLKSILVEDEDAEGKVLFTDEIHLVLAGKTEGMD 295
 237 RKLISLDMGALVAGAKYRGEFEERLKSILVEDEDAEGKVLFTDEIHLVLAGKTEGMD 296
 296 AANLKPMLARGQLRCIGATTLEBYRYKYVEKDAFERFQVVAEPVDTTISILGLK 355
 297 AGNLLKPMARGQLRCIGATTLEBYRYKYVEKDAFERFQVVAEPVDTTISILGLK 356
 356 EKYEGHGVRIQDRLINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPBEIND 415

357 ERYEVHGVKADSAALVAAMLSNRYISDRFLPKAIDLVDVDAKAKLMEITSPKPELDE 416
 416 LERKEMOLEIEHLEHAREKOKASKARLIEVKEKELDLKLOPLTKYRKEKERIDIR 475
 417 VDRKILQEMERLSLQRENDASERLEKLEKELADFKKEOSKLNQWQSEKTVIDQIRT 476
 476 LKQREELMFLSLOAERRYDLARAADLYGAIQVESAIAOEGTSSEENV---MLTEN 531
 477 VKETIDVNLBEIQQAQRDYDKAAELQYKGLTDLQVQVEALEFQLAEQOTSGKSLREE 536
 532 VGPHEIAEVRWRTGPIVTRIGQNEKERLIGLARLHKRVVQGNQAVNAVEAILRSRAG 591
 537 VLESIDIAEIIKWTGIPISKVASEKELHLEDELHRSRVIGQDEAVTAVAEAIQSRAG 596
 592 LGRAQOPTGSLFGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVRLIGAP 651
 597 LSDNRPFTAFIFGPTGVGKTELAKALAKNLFTTEALVRIDMSEYMEKHAVSRIMGAP 656
 652 PGYVHGEGGQTEAVRRPVCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFR 711
 657 PGYVHGEGGQTEAIRRPVSILFDEIEKAGGVFNWMLQILDGRLTDGQRTVDFR 716
 712 NSVIMTSLNLAELHLAGLTKVMEVARDVMEVREKHFPELNLRLDRIVDFPLSHD 771
 717 NTIIMTSLNLSQVILVDAGDSDRYEEMSRVMDVMEFPEFNRVDETIIFHGLQKS 776
 772 QLRKVARLQMKDVAVRLAERVALAVTDALDYLAESYDPVYVGARPIREMEKVVTEL 831
 777 ELSRIVQIQSLATREBEQKLTUKTKALDFLAAGVDPVYVGARPLKRAVOKYLETAI 836
 832 SKMVREIDENSTVYIDAGAGDLVYRVESGGLV 865
 837 AKGILRGDYKPGETIVVDDETDERLSFTSLRGDLV 870

JUL 15

49337

ABB49337 standard; Protein; 866 AA.

05-FEB-2002 (first entry)

Listeria monocytogenes protein #2041.

Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation; vitamin B12; bacterial infection; disease.

Listeria monocytogenes.

WO200177335-A2.

18-OCT-2001.

11-APR-2001; 2001WO-FR01118.

11-APR-2000; 2000FR-0004629.

(INSP) INST PASTEUR.

Ruchrieser C, Prangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P; Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P; Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA; Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A; Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L; Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N; Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J; Rose M, Voss H;

WPI; 2002-010914/01.

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

and prevention of Listeria and related bacterial infections, and related polypeptides

Claim 6; SEQ ID No 2042; 192pp; French.

The present invention relates to the genome sequence of Listeria monocytogenes BGD-e (see ABA03041). The genome sequence and fragments of it are useful for selecting probes and primers for detecting genes in L. monocytogenes and related organisms, and for studying genetic polymorphisms and other genomes. The present sequence is a protein encoded by the genome sequence of the present invention. Proteins expressed from the genome sequence are useful for raising specific antibodies. Identification of L. monocytogenes and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate L. monocytogenes-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccine compositions for the treatment or prevention of infections by L. monocytogenes and related organisms.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPD at ftp.wipo.int/pub/published_pct_sequences.

Sequence 866 AA;

Query Match 49.2%; Score 2236.5; DB 23; Length 866;
 Best Local Similarity 51.5%; Pred. No. 8.9e-158;
 Matches 448; Conservative 181; Mismatches 224; Indels 17; Gaps 9;

QY 1 MNPEKFTKNTETATATHELVNAGHAQFTPLHLAALISDPTGIFPQAISAGGENAAQ 60
 DB 1 MDLQFTQVQQTADAQNLAIASEHQIDVAVPKVLLTSED--FAKRVYDV-AEVDTD 57
 QY 61 SAERVINGALKLP--SQSPPPDDIPASSSTIKVIRRAQAQKSGDTHLAVDQLIMGL 118
 DB 58 ALQKVIENTLEKIPVVGSGVNYGQMSQALFQLMRDAKEQQLQEDDFVSTEHLILAV 117
 QY 119 E--DSQIRDLNEVGAVATARKSEVKELRGKGVESASGDTNFOALKTYGRDLVEQ-- 174
 DB 118 DQKSNPITAEKQKAKKQKAEILKIRG--GKRVTSQNAENYEALTKYGRDLVAEVR 175
 QY 175 AGKLDPVIGRDEIRRVVRLISRTKNNPVLIGPGVGTAVVEGLAQRIYKGVDPNSLT 234
 DB 176 SKGLDPVIGRDAEIRNVIRLSRTKNNPVLIGPGVGTAVVEGLAQRIYKGVDPNSLT 235
 QY 235 DVRLISLDMGALVAGAKYRGFEERLKSVLKEVDBAGKVLFTDEIHLVLGACKTSGM 294
 DB 236 DXTIISLDIGSLIAGAKYRGFEERLKSVLKEVDBAGKVLFTDEIHLVLGACKTSGM 295
 QY 295 DAANLFKPMALARGQLRCIGATTLSEYKYVEKDAAFRRFPQVYVAPSVDDTISIRGL 354
 DB 296 DAGNMLKPMALARGELHCIGATTLSEYKYVEKDAAFRRFPQVYVAPSVDDTISIRGL 355
 QY 355 KEKYEKGHVRIQDRALINAAQLSARYITGTHLDPKALDVLDEACANVRVQDSQPEID 414
 DB 356 KERFEIHHGVNIHDNALVAASLSNRVITDRLPDKALDVLDEACATIRVEIDMPSELD 415
 QY 415 NLERKMOLEIELHALEREKOKASKARLIEVRKELDDLRLDKLOPLTKYRKEKERIDEIR 474
 DB 416 EVTRKVMQLETEAAKKEKDPASERLEILQRLADYKBEANQMSKSEKNEISKIR 475
 QY 475 RLKQKRELMESLOEASRRYDLAADAIRLGAIOVESATAQLEGTSSE----ENVMLTE 530
 DB 476 EVRFQIDHLRHELESEANNVNLKAAELRHGRIPAVEKELLELEAEENREKTAQEDRLQ 535
 QY 531 NVGPEHIAEVVSRWTGIPVTRIGQNEKERLIGLARLHKRVVQGNQAVNAVEAILRSRA 590
 DB 536 EVTENEIAEIVGRWTGIPVTKLVEGEREKLLKADVLHVKVIGQDDAVQLVDAVLARA 595
 QY 591 GLGKRAQOPTGSLFGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVRLIGA 650
 DB 596 GIKDKPKRPIGSGFIFLPGTGVGKTELAKALAFNMFDSSEHDMIRIDMSEYMEKHSVRLIGA 655

651 PPGYVGEHGGQTEAVRRPYPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDF 710
 656 PPGYVGEHGGQTEAVRRPYPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDF 715
 711 RNSVIIMTSLGAEHLLAGL-TGKVTMEVARDVREVRKHFRPELNLRLDEIVVDPPLS 769
 716 KNTVIIMTSLGAEHLLAGL-TGKVTMEVARDVREVRKHFRPELNLRLDEIVVDPPLS 774
 770 HDQLRKVARLQMKDVAVRLAERGVVALAVTDAAALDYILARSYDPVYGARPIRMEKKVVT 829
 775 LADIKGIVEKLVELOIRLADQBITTISDAAKAFIABEAYDPVYGARPIRMEKKVVT 834
 830 ELKMWVRBEIDENSTVYIDAGADLVYRV 859
 835 PLAREIVSGKIMPHSSVEIDLADKEFTKV 864

arch completed: February 13, 2004, 01:30:55
 o time : 77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 13, 2004, 01:10:25 ; Search time 47 seconds
(without alignments)
1864.036 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

quence: 1 MNPEKFTHTNETIATAHEL.....VKOMRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 283308 seqs, 96168682 residues

al number of hits satisfying chosen parameters: 283308

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult	No.	Score	Query Match	Length	DB ID	Description
1	4545	99.9	911	2	F96771	heat shock protein
2	4003	88.0	911	1	T07807	endopeptidase Clp
3	2807	61.7	831	2	D71409	probable endopepti
4	2431	53.4	668	2	G85160	heat shock protein
5	2294	50.4	931	2	AF3276	ATP-dependent clp
6	2288	50.3	872	2	S76197	endopeptidase Clp
7	2258.5	49.6	865	2	D97018	ATPase with chaper
8	2242.5	49.3	866	2	A11720	endopeptidase Clp
9	2236.5	49.2	866	2	AF1250	endopeptidase Clp
10	2234.5	49.1	905	2	T39572	probable proteinase
11	2224	48.9	859	2	C87358	hypothetical prote
12	2218.5	48.8	835	2	AD2441	endopeptidase Clp
13	2212.5	48.6	862	1	E35905	endopeptidase Clp
14	2204	48.4	857	2	AI0831	ClpB protein (heat
15	2199	48.3	874	2	AC3070	ATP-dependent Clp
16	2199	48.3	887	2	F98216	endopeptidase Clp
17	2193	48.2	857	1	D35905	endopeptidase Clp
18	2193	48.2	857	2	G91060	heat shock protein
19	2193	48.2	861	2	E85505	heat shock protein
20	2182.5	48.0	898	2	S76431	endopeptidase Clp
21	2168.5	47.7	858	2	D71711	endopeptidase Clp
22	2166	47.6	968	2	T51523	clpB heat shock pr
23	2163.5	47.5	857	2	C97707	clpB protein (impo
24	2162.5	47.5	856	1	F64098	endopeptidase Clp
25	2160	47.4	857	2	AI0397	Clp ATPase [import
26	2158.5	47.4	848	2	C70834	probable endopepti
27	2158.5	47.4	861	2	D82814	ATP-dependent Clp
28	2153.5	47.3	853	2	T36551	probable ATP-depen
29	2150.5	47.3	848	2	G87220	heat shock protein

30	2141.5	47.1	869	2	H89864	hypothetical prote
31	2133.5	46.9	880	2	AC2096	endopeptidase Clp
32	2133	46.9	854	2	D83077	ClpB protein PA454
33	2119.5	46.6	875	2	G75442	ATP-dependent Clp
34	2104.5	46.3	859	2	F81078	clpB protein NMB14
35	2104	46.2	878	2	G71371	probable endopepti
36	2103.5	46.2	859	2	F81863	clpB protein NMA16
37	2099.5	46.1	857	2	E86815	clpB protein (impo
38	2086	45.8	857	2	A82290	clpB protein VC071
39	2069	45.5	874	2	G84644	probable ATP-depen
40	2039	44.8	860	1	C35905	endopeptidase Clp
41	1943	42.7	867	2	A81707	ATP-dependent Clp
42	1938	42.6	867	2	D71555	endopeptidase Clp
43	1927.5	42.4	857	2	F81396	ATP-dependent Clp
44	1912	42.0	866	2	E72113	clp proteinase ATP
45	1912	42.0	866	2	H86508	clp proteinase ATP

ALIGNMENTS

RESULT 1

F96771

heat shock protein 101, 13093-16240 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C;Accession: F96771

R;Theologias, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso, ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Chin, C.W.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maili, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96771

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-911 <STO>

A;Cross-references: GB:AE005173; NID:G6939223; PIDN:AAF31725.1; GSPDB:GN00141

C;Genetics:

A;Gene: F1017.2

A;Map position: 1

C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 99.9%; Score 4545; DB 2; Length 911;
Best Local Similarity 99.9%; Pred. No. 6.6e-194;
Matches 910; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MNPEKFTHTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISSAGGENAAQ	60
Db	1	MNPEKFTHTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISSAGGENAAQ	60
QY	61	SAERVINALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGDTPLAVDQIMGLLED	120
Db	61	SAERVINALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGDTPLAVDQIMGLLED	120
QY	121	SGIRDLLNEGVATARKVSEVEKLGRGKGVESAGDTNFAQKTYGRDLVEQAGKLDLP	180
Db	121	SGIRDLLNEGVATARKVSEVEKLGRGKGVESAGDTNFAQKTYGRDLVEQAGKLDLP	180
QY	181	VIGRDEEIRRVRLISRRNTKPNVLIGEPGVGKTAVVEGLAQRIKVGSDVNSLTDVRLIS	240
Db	181	VIGRDEEIRRVRLISRRNTKPNVLIGEPGVGKTAVVEGLAQRIKVGSDVNSLTDVRLIS	240
QY	241	LDMGALVAGAKYRGFEFERLKSVLKEVEDARGKVLFDITDIHLVLGAGKTEGSDAANLF	300
Db	241	LDMGALVAGAKYRGFEFERLKSVLKEVEDARGKVLFDITDIHLVLGAGKTEGSDAANLF	300
QY	301	KPMLARGQLRCIGATTLEERYKVEKDAAFERFPQVVVAEPSPVDTISILRLGLKEKYE	360

RESULT 3
D71409
probable
N; Alternat
N; Contain
C; Species

ariety: columbia
 Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 19-Jan-2001
 Accession: D71409
 Medler, H.; Wedler, E.; Kambutt, T.; Pohl, T.M.; Terry, N.; Giel
 agn, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 are 391, 485-488, 1998
 Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
 ft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; And
 Chaltawitz, N.
 tle: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 ference number: A71400; MUID:98121113; PMID:9461215
 Accession: D71409
 Status: preliminary; nucleic acid sequence not shown; translation not shown
 Molecule type: DNA
 Residues: 1-831 <BV>
 Cross-references: GB:297336; NID:G2244788; PIDN:CAB10246.1; PID:G2244823
 Genes:
 Map position: 4COP9-4G3845
 Function:
 Description: allows clip to hydrolyze polypeptides and proteins, probably by a chaperon
 superfamily: endopeptidase Clp ATP-binding chain
 Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop
 12-179/Region: nucleotide-binding motif A (P-loop)
 10-245/Region: nucleotide-binding motif B
 11-578/Region: nucleotide-binding motif A (P-loop)
 0-645/Region: nucleotide-binding motif B
 8/Binding site: ATP (Lys) #status predicted
 7/Binding site: ATP (Lys) #status predicted
 Query Match 61.7%; Score 2807; DB 2; Length 831;
 Best Local Similarity 64.1%; Pred. No. 5e-117;
 Matches 585; Conservative 108; Mismatches 129; Indels 90; Gaps 11;
 1 MNPEKFTKNTETATATHELAVNAGHAQFTPLHLAALISDPTGIPFOAISSAG-GENAA 59
 1 MNDLKFDPNVKLILASARSAMSLSHGQVTPHLGVTLISDLTSVFYRAITSAGDGDISA 60
 60 QSAERVINQALKPLSQSPPPDDIPASSSLIKVIRRAQAQKSRGDTPLAVDQIMGLLE 119
 61 QSVVNVINQSLYL-----TKRNLGDTKGVAVLVISLLE 95
 120 DSQIRDLNNEVGAVATARKVSEKLRGKGGKVESASGDTNFQALKTYGRDLVEQAGKLD 179
 96 DSQISDLVKEAGVVPKSEVEKL-----GEVILRALTKYGTDLVEQAGKLD 144
 180 PVIGRDEEIRRVIRILSRRTKNNPVLIGPVGKTA VEGLAQRIKGVDPNSLTDVRLI 239
 145 PVIGRDEEIRRVIRILSRRTKNNPVLIGPVGKTA VEGLAQRIKGVDPINLTGVKLI 204
 240 SLDMGALVAGAKYGRFEERLKSVLKEVEDAEQKVLIFDEIHLVLAGKTEGSDAANL 299
 205 SLEFGAMVAGTTLRGQFEERLKSVLKAVEAQKVLVFIDEIHLVLAGKACASGSTDAKL 264
 300 FKPMALRGQLRCIGATTLEBYRYKVEKDAAFERRFOQVYVABSPVDDTISILRLKKEYE 359
 265 LKPMALRGQLRCIGATTLEBYRTHVEKDAAFERRFOQVYVABSPVDDTISILRLKKEYE 324
 360 GHGVRIQDRALNNAQLSARYITGRLHPDKAIDLVDACANVRVQLDSQPEINLERK 419
 325 GHGVRIQDRALVLSAQLSERYITGRLHPDKAIDLVDSCAHVKAQLDQPEIDSLERK 384
 420 RMQLEIHLALEREK-DKASKARLIEVRKELDLRLKQLPMYRKKEKERIDEIRRLKQ 478
 385 VMQLEIHLALEKEDKASEARLSEVRKELDLRLKLEPLIKYKKEKKIINETRLKQ 444
 479 KREELMFSIQEARRRYDLARADRYGAIOEVESAIAQLEGTSSEENVMLTENVGPEHIA 538
 445 NRDDLMLAQEARRQHDVPAKAVLYKYGAIOEVESAIAKLE-KSADKNVMLTETVGPENIA 503
 539 EVVSRWTGIPVTRLDQNEKKRLISLADKLHERVVGQDEAVKAVAAAILRSRVGLGRPOOP 563

Db QY 504 EVVSRWTGIPVTRLDQNEKKRLISLADKLHERVVGQDEAVKAVAAAILRSRVGLGRPOOP 563
 QY 599 TGSFPLFGTGVGKTELAKALAEQFDENLVRIDMGEYMEQHSVRLIGAPPG-YVGH 657
 Db 564 SGSFPLFGTGVGKTELAKALAEQFDENLVRIDMGEYMEQHSVRLIGAPPGYIGH 623
 QY 658 EGGGOLTRAVRRPVCVILFDEVEKARHVAVENTLLQVLDGDLTQCGRTVDFRNSVLIIM 717
 Db 624 EGGGOLTRAVRRPVCVILFDEVEKARHVAVENTLLQVLDGDLTQCGRTVDFRNSVLIIM 673
 QY 718 TSNLGAHLLAGLTKGVMEVARDCVMEVRKHFPELNLRIIDEIVDFPLSHDQLRKVA 777
 Db 674 TSNLGAHLLVSLTCEMTQVARDNAMDKKHFPELNLRIIDEIVDFPLSHDLAKIV 733
 QY 778 RLOMKOVAVRLAERGVALLVDAALDYILAESYDPVYGARPIRMMEKKVVTLSKVVVR 837
 Db 734 QLOQKXNV-----PIRWLERKVVTDISMIVR 761
 QY 838 REIDENSTVIYDAGAG--DLVVRVESGGLVDASTGKSDVLIHIANGPKRSDAAQAV--- 892
 Db 762 BEIGDSDVICIDVNEAKTDLVYRDKNVFVKIE--QTLVDVVHSGNKGKGRNDEHDHIVTL 819
 QY 893 -KCOMRIEIEDD 903
 Db 820 TKKIKNEVVVID 831
 RESULT 4
 G85160
 heat shock protein like [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
 C:Accession: G85160
 R:anonymous; The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
 A:Reference number: A85001; MUID:20083488; PMID:10617198
 A:Accession: G85160
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-868 <STO>
 A:Cross-references: GB:NC_001268; NID:G5302773; PIDN:CAB46061.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: dl3375w
 A:Map position: 4
 C:Superfamily: endopeptidase Clp ATP-binding chain
 Query Match 53.4%; Score 2431; DB 2; Length 668;
 Best Local Similarity 70.8%; Pred. No. 1.7e-100;
 Matches 499; Conservative 76; Mismatches 90; Indels 40; Gaps 6;
 QY 1 MNPEKFTKNTETATATHELAVNAGHAQFTPLHLAALISDPTGIPFOAISSAG-GENAA 59
 Db 1 MNDLKFDPNVKLILASARSAMSLSHGQVTPHLGVTLISDLTSVFYRAITSAGDGDISA 60
 QY 60 QSAERVINQALKPLSQSPPPDDIPASSSLIKVIRRAQAQKSRGDTPLAVDQIMGLLE 119
 Db 61 QSVVNVINQSLYL-----TKRNLGDTKGVAVLVISLLE 95
 QY 120 DSQIRDLNNEVGAVATARKVSEKLRGKGGKVESASGDTNFQALKTYGRDLVEQAGKLD 179
 Db 96 DSQISDLVKEAGVVPKSEVEKL-----GEVILRALTKYGTDLVEQAGKLD 144
 QY 180 PVIGRDEEIRRVIRILSRRTKNNPVLIGPVGKTA VEGLAQRIKGVDPNSLTDVRLI 239
 Db 145 PVIGRDEEIRRVIRILSRRTKNNPVLIGPVGKTA VEGLAQRIKGVDPINLTGVKLI 204
 QY 240 SLDMGALVAGAKYGRFEERLKSVLKEVEDAEQKVLIFDEIHLVLAGKTEGSDAANL 299
 Db 205 SLEFGAMVAGTTLRGQFEERLKSVLKAVEAQKVLVFIDEIHLVLAGKACASGSTDAKL 264
 QY 300 FKPMALRGQLRCIGATTLEBYRYKVEKDAAFERRFOQVYVABSPVDDTISILRLKKEYE 359

[illegible]

RESULT 6
 S76197
 endopeptidase Clp ATP-binding chain B1 - *Synechocystis* sp. (strain PCC 6803)
 N:Alternate names: ATP-dependent Clp proteinase regulatory chain; protein sir1641
 N:Contains: adenosinetriphosphatase (EC 3.6.1.3)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 19-Jan-2001
 C:Accession: S76197
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76197
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-872 <AN>
 A:Cross-references: EMBL:D90914; GB:AB001339; NID:gl653477; PIDN:BAAL8456.1; PID:gl65354
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: clpB1
 C:Function:
 C:Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon-
 e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
 C:Superfamily: endopeptidase Clp ATP-binding chain
 C:Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop
 F:208-215/Region: nucleotide-binding motif A (P-loop)
 F:276-280/Region: nucleotide-binding motif B
 F:611-618/Region: nucleotide-binding motif A (P-loop)
 F:679-683/Region: nucleotide-binding motif B


```

777 EIGKTIIDIFLENKSKLEKNTKIDIAEAKKIIAEEGYDPVYGARPLKRYTIENTIETHI 836
832 SKWVVEEIDENSTVYIDAGDGL 855
837 AKMFTSGEISEGDIKIGSDSKL 860

RESULT 8
720
endopeptidase Clp ATP-binding chain B (ClpB) homolog clpB [imported] - Listeria innocua
Species: Listeria innocua
Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
Accession: A11720
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
Title: Comparative genomics of Listeria species.
Reference number: AB1077; MUID:21537279; PMID:11679669
Accession: A11720
Status: preliminary
Molecule type: DNA
Residues: 1-866 <GLA>
Cross-references: GB:NC_003210; PID:CAD00284.1; PID:gl6411676; GSPDB:GN00177
Experimental source: strain Clp11262
Genetics:
A:Gene: clpB
Superfamily: endopeptidase Clp ATP-binding chain

Query Match 49.3%; Score 2242.5; DB 2; Length 866;
Best Local Similarity 51.6%; Pred. No. 5.1e-92;
Matches 449; Conservative 185; Mismatches 219; Indels 17; Gaps 9;

1 MNPEKTHKTNETIATAHELAVNAGHAQTPHLAGALISDPTGIFPOAISAGGENAAQ 60
1 MDLQKFTQVQQTIAQAONLATASEHQEIDVAHFVKVLTESD--FAKRVYDV-AEVD 57
61 SAERVINOALKKLP--SQSPPPDDIPASSSLIKVIRRAQAQSRGDTHLAVDQIMGLL 118
58 ALQKVVDSELRKIPVVGSGVNGVQMSQALFQMRDEKEQQLDDFVSTHLLILAVM 117
119 E--DSQIRDLNNEGVATARVSEVKEKLGKGGKVESASGDTNFGALKTYGRDIVEQ-- 174
118 DQKSNPITVELKQNKQSKQINEAILKIRG--GKRVTSQNAEENYEALTKYGRDLVAVR 175
175 AGKLDPVIGRDEIRRVVRLSRRTKNNPVLIGERGKGTAVVEGLAQRIYKGVDPNSLT 234
176 SKLDPVIGRDAEIRNVIRILSRKTKNNPVLIGEPGVGKTAIVEGLAQRIYKGVDPNSLT 235
235 DVRLISLDMGALVAGAKYGEPEERLKSYLEVEDEAGKVIILFIDEIHLVLCAGKTEGSM 294
236 DKTIIISLIGSLIAGAKYGEPEERLKAFLQVQSDGQILLFIDEIHTIVGAKTGDAM 295
295 DAANLFPKMLARGQLRCIGATTLEIRYKVEKDAAFERRPQOVVVAEVPDPTISILRGL 354
296 DAGNMLKPMALRGELHCIGATTLEIRYQVIEKDAALERRFQKVLVPEPTVEDTVSILRGL 355
355 KEKYEGHGVRIQDRALINAAQLSARYITGRLHPKADLDLVEACANVRQLDSQPEID 414
356 KERFEIHGVNTHNALVAASLSNRYITDRFLPKADLDLVEACATIRVDSNPSELD 415
415 NLERKMQLEIHALEREKQKASKARLIEVRKELDLRLDKLPITMKYRKEKSIDIR 474
416 EVTRKVMQLEIEAALKEKDPASRRLEMLQRELADYKEANKMKSWSEKSEISKIR 475
475 RLKQKRELMFSLQEAERYDLARAADLYRGAIQEVESAIQLETSSE----ENVMLTE 530
476 EVREQIDHLRHELEEAENNYDLNKAELRHGKIPAVEKELLALETENREKTAQDRILQ 535
531 NVGPEHIAEVSWSRTGIPVTRIGQNEKERLIGLADRLHKRIVGVQNAVNAVSEALRSRA 590

```

```

536 EVTENEAEIVGRTGIPVTKLVEGEREKLLKLVADVLHOKVIGQDDAVQLVSDAVLRARA 595
591 GLGRAQOQPTSGFLPGPTGVGKTELAKALAEQIFDDENLLVRIDMSEYMEQHSVRLIGA 650
596 GKDKPKPIGSIPLGPTGVGKTELAKALAYNNFDSHMRIDNSEYMEKHSVRLVGA 655
651 PPGYVGHGEGQLTEAVRRRPYCVILFDEVEKAHVAVENTLLOVLDGGRITDCQGRTVDF 710
656 PPGYVGHGEGQLTEAVRRRPYCVILFDEVEKAHVAVENTLLOVLDGGRITDCQGRITDF 715
711 RNSVIIMTSMGAEHLAGL--TGKVTMEVARDCVMEVRKHFPELNLRLDEIVFPDPLS 769
716 KNTVIIMTSMGAEHLAGL--TGKVTMEVARDCVMEVRKHFPELNLRLDEIVFPDPLS 774
770 HDQLRKVARLQMDKVAVRGAERGVAVLADTADLADYILASYPDVGARPIRMWKKVVT 829
775 LADIKGIVEKLELQRLADQBITITISDNKAFABEAYDPVYGARPLKRYIVRHVET 834
830 ELSKMVVEEIDENSTVYIDAGDGLVYRV 859
835 PLAREIVSGKIMPHSSVEIDLQDKETFKV 864

RESULT 9
AF1350
endopeptidase Clp ATP-binding chain B (ClpB) homolog clpB [imported] - Listeria monocytogenes
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
Accession: AF1350
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
Accession: AF1350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-866 <GLA>
A:Cross-references: GB:NC_003210; PID:CAD00284.1; PID:gl6411676; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: clpB
C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 49.2%; Score 2236.5; DB 2; Length 866;
Best Local Similarity 51.5%; Pred. No. 9.5e-92;
Matches 448; Conservative 181; Mismatches 224; Indels 17; Gaps 9;

1 MNPEKTHKTNETIATAHELAVNAGHAQTPHLAGALISDPTGIFPOAISAGGENAAQ 60
1 MDLQKFTQVQQTIAQAONLATASEHQEIDVAHFVKVLTESD--FAKRVYDV-AEVD 57
61 SAERVINOALKKLP--SQSPPPDDIPASSSLIKVIRRAQAQSRGDTHLAVDQIMGLL 118
58 ALQKVIENTLEKIPVVGSGVNGVQMSQALFQMRDEKEQQLDDFVSTHLLILAVM 117
119 E--DSQIRDLNNEGVATARVSEVKEKLGKGGKVESASGDTNFGALKTYGRDIVEQ-- 174
118 DQKSNPITVELKQNKQSKQINEAILKIRG--GKRVTSQNAEENYEALTKYGRDLVAVR 175
175 AGKLDPVIGRDEIRRVVRLSRRTKNNPVLIGERGKGTAVVEGLAQRIYKGVDPNSLT 234
176 SKLDPVIGRDAEIRNVIRILSRKTKNNPVLIGEPGVGKTAIVEGLAQRIYKGVDPNSLT 235
235 DVRLISLDMGALVAGAKYGEPEERLKSYLEVEDEAGKVIILFIDEIHLVLCAGKTEGSM 294
236 DKTIIISLIGSLIAGAKYGEPEERLKAFLQVQSDGQILLFIDEIHTIVGAKTGDAM 295
295 DAANLFPKMLARGQLRCIGATTLEIRYKVEKDAAFERRPQOVVVAEVPDPTISILRGL 354
296 DAGNMLKPMALRGELHCIGATTLEIRYQVIEKDAALERRFQKVLVPEPTVEDTVSILRGL 355
355 KEKYEGHGVRIQDRALINAAQLSARYITGRLHPKADLDLVEACANVRQLDSQPEID 414
356 KERFEIHGVNTHNALVAASLSNRYITDRFLPKADLDLVEACATIRVDSNPSELD 415
415 NLERKMQLEIHALEREKQKASKARLIEVRKELDLRLDKLPITMKYRKEKSIDIR 474
416 EVTRKVMQLEIEAALKEKDPASRRLEMLQRELADYKEANKMKSWSEKSEISKIR 475
475 RLKQKRELMFSLQEAERYDLARAADLYRGAIQEVESAIQLETSSE----ENVMLTE 530
476 EVREQIDHLRHELEEAENNYDLNKAELRHGKIPAVEKELLALETENREKTAQDRILQ 535
531 NVGPEHIAEVSWSRTGIPVTRIGQNEKERLIGLADRLHKRIVGVQNAVNAVSEALRSRA 590

```

296 DAGNNLKEMLARGELHCIGATTLEQYVIEKDALERRFQKVLVPEPTVEDTVSILRGL 355
 355 KEKVEGHEGVRIQDRALNAALQASRYITGRHLPDKALDLVDEACANVRVOLDQSOPEID 414
 356 KERFEIHGVNHVHVALVAASLSNRYITDRFLPDKALDLVDEACATIRVEIDSNPSELD 415
 415 NLERKMOLEIEIHALEREKOKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIR 474
 416 EVTRKVMQLEIEEAAKEKOPASERRLEIIORELADYKBEANQMKWSEKNEISKIR 475
 475 PLKQKREBELMESLOEAERRYDLARAADLYRGAIOEVESAIAOEGTSSE-----BNVMLTE 530
 476 EVREQIDHRLHELEEAENNYDLNKAALRHGRIPAVEKELLELEAENREKTAQADRILQE 535
 531 NVGPEHIAEVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVVGQONAVNVAISRA 590
 536 EVTENEAIEIVGRWTGIPVTLKVEGEREKLKLDVHLHQVIGQDDAVQLVSDAVLRARA 595
 591 GLGQAQPTGFLGPTGVGKTELAKALAQFLPDDENLLVRIDMSYMEQHSVSRILGA 650
 596 GIKDKPRIGSIFLGPVGKTELAKALAFNMFDSBDHMIRIDMSYMEKHSVSRILGA 655
 651 PPGVYVGEHGGQLTEAVRRRYPYCVILDEVEKAHVAVNTLLQVLDGRLTDGGRVTDF 710
 656 PPGVYVGEHGGQLTEAVRRRYPYCVILDEVEKAHPDVNTLLQVLDGRLTDGGRVIDF 715
 711 RNSVIMTSNLSGASHLAGL--TGKVTMEVARDCVMRVRRKHFRPELLNRLDEIVDFPLS 769
 716 KNTVIMTSNLSGALLERTEEGISEPELSD--VMQIQSEFKPEFLNRVDDILFLKPLT 774
 770 HDQURKVARLQMDKAVRLAERGVALVTDAAALDYILAESYDPVYGARPIRRMWEKKVVT 829
 775 LADIKGIVEKLVEQLRLAQDEITITISDDAKAFIAEEAYDPVYGARPLKRYIVRHVET 834
 830 ELSRWVREBEIDENSTYIIDAGADLYVRV 859
 835 PLAREIVSGKIMPHSSVEIDLADKEFTFKV 864

JUL 10
 772
 able proteinase subunit - fission yeast (Schizosaccharomyces pombe)
 Species: Schizosaccharomyces pombe
 Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Feb-2001
 Accession: T39572
 Codon, V.; Rajandream, M.A.; Bartell, B.G.; Brown, D.; Churcher, C.M.
 mitted to the EMBL Data Library, March 1999
 reference number: 221864
 Accession: T39572
 Status: preliminary; translated from GB/EMBL/DDBJ
 molecule type: DNA
 residues: 1-905 <WOO>
 cross-references: EMBL:AL035637; PIN: CAB38512.1; GSPDB: GNO0067; SPDB: SPBCL6D10.08c
 experimental source: strain 972h-; cosmid c16D10
 netics:
 Gene: SPDB: SPBCL6D10.08c
 ap position: 2
 Superfamily: endopeptidase Clip ATP-binding chain
 Keywords: ATP; molecular chaperone; nucleotide binding; P-loop
 19-216/Region: nucleotide-binding motif A (P-loop)
 77-282/Region: nucleotide-binding motif B
 21-628/Region: nucleotide-binding motif A (P-loop)
 99-694/Region: nucleotide-binding motif B
 15/Binding site: ATP (Lys) #status predicted
 27/Binding site: ATP (Lys) #status predicted

ery March 49.1%; Score 2234.5; DB 2; Length 905;
 est Local Similarity 49.5%; Pred. No. 1.2e-91;
 atches 460; Conservative 158; Mismatches 258; Indels 53; Gaps 11;
 6 FTHKNTNTIATAHELVNAGHAQPTPLHLAGALISDP----TGIFPQAITSAGGEVNAQS 61

Db 6 FTDKAATLSDAYSIAQSYGHSQLTPHIAAALLSDSDSGTTLRLTVDKAGD--GQK 63
 QY 62 AERVINQALKKLPSQPPDDIPASSLLIKVIRRAQAOKSRGDTHLAVDQLIMGLLEDS 121
 Db 64 FERSVTSRLVRUPAOPPPPEQVTLSPESAKLRLNAHELQTKQKSYIAQDHFIAVFTKDD 123
 QY 122 QIRDLNEVGVATARVKSVEKLGKGEKGVESASGDTNFQALKTYGRDLVBEQA--GKLD 179
 Db 124 TLKSLLAEGAVTPKAFEPANVRG--NKRIDSNAEBGFDALNKFTVDLTTELARGQLD 181
 QY 180 PVIGRDEETRRVVRILSRRTKKNPVLIGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 Db 182 PVIGRDEETRRVVRILSRRTKKNPVLIGEPGVGKTAVVEGLARRIIDDDVPANLSCKLL 241
 QY 240 SLDMGALVAGKVRGEFERELKSVLKEVEDAEKGVILFIDEHLVLGAK--TEGSMDAAN 298
 Db 242 SLDVGLVAGSRGEFEERIKSVLKEVESESTPILFVDEMHLMGAGSGGEGMDAAN 301
 QY 299 LFKPMLARGQLRCIGATTLLEERYKRVKEDAAFERRFQVYVAEPSPVDTISILRGLEKY 358
 Db 302 LKPLMLARGKLHCIGATTLAEYKVIKEDAAFERAFQILVKEPSIEDTISILRGLEKY 361
 QY 359 EGHGVRIODRALINAAQLSARYITGRHLPDKAIDLVEACANVRVOLDQSOPEIDNLER 418
 Db 362 EVHGVTTISDRALVTAHLASRYLTSRLPDSALDLVDEAAAARVTVRESOPEVLDNLER 421
 QY 419 KRMOLIEIHALEREKOKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRRLKQ 478
 Db 422 KLRQURVEITRALERBDEASKERLKAARKEAQVEETRPTRKEYELEKRSGELODAKR 481
 QY 479 KREELMFSLOEAERRYDLARAADLYRGAIOEVESAIAOEGTSSEENV-----526
 Db 482 RLDELKAKAEADAERNDFTLAADLYKVPDLQKRIEYLEQOKRKADAEATANAQPSGP 541
 QY 527 MLTENVGPEHIAEVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVVGQONAVNVAIS 586
 Db 542 LLIDVVGPDQINEIVARWTGIPVTRLKTTKEKRLNMEKVLKQVIGQNEAVTAVANAI 601
 QY 587 RSRAGLGRAQPTGSGFLGPTGVGKTELAKALAQFLPDDENLLVRIDMSYMEQHSVSR 646
 Db 602 LSRAGLSDPNQPIASFLGPGSGTGKTLTKALASFMFDDENAMIRIDMSYMEKHSVSR 661
 QY 647 LIGAPPVYVGEHGGQLTEAVRRRYPYCVILDEVEKAHVAVNTLLQVLDGRLTDGGR 706
 Db 662 LIGAPPVYVGEHGGQLTEQLRRRYPYCVILDEVEKAHPEVLTVLLQVLDGRLTDGGR 721
 QY 707 TVDFNSVIMTSNLSGASHILA--GLTGKVTMEVARDCVMRVRRKHFRPELLNRLDEIV 763
 Db 722 VVDKNAVIMTSNLSGASHILATDNESSDDGKID--STTREMVMNSIRGFFRPFELNRISIV 780
 QY 764 VFDPLSHDQLRKVARLQMDKAVRLA--ERGVALAVTDAALDYILAESYDPVYGARPIRR 821
 Db 781 IFNRLRRVDIRNIVENRILEVQKRLQSNRSIKTEVSEADKDLGSGAGSYGARGPLNR 840
 QY 822 WMEKKVTVTELSRWVREBEIDENSTYIIDAGADLYVRVRESGGLVDASTGKKSVDLIHTAN 881
 Db 841 VTQNGVLNPMVAVLINGQLRKETAHVVVVQNGKIFVKPN-----HEAN 883
 QY 882 GPKRSDAAQVKKMKRIEIDDDNEEMIE 910
 Db 884 ANGSAD-----IDMDGIDDDVDNDELE 905

RESULT 11

C87358

hypothetical protein C0878 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: C87358

R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

Title: Complete Genome Sequence of Caulobacter crescentus.

Reference number: AB7249; MUID:21173698; PMID:11259647

Accession: C87358

Status: preliminary

Molecule type: DNA

Residues: 1-859 <STO>

Cross-references: GB:AE005673; NID:g13422139; PIDN:AAK22863.1; GSPDB:GN00148

Genetics:

Gene: CC0878

Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.9%; Score 2224; DB 2; Length 859;

Best Local Similarity 51.9%; Pred. No. 3.4e-91;

Matches 446; Conservative 158; Mismatches 246; Indels 10; Gaps 6;

1 MNPEKFTHTNNTIATAHELVNAGHAQPTPLHLAGALISDPTGTFPPQAISSAGGENAAQ 60

1 MNIDLYSDKAKAVASQASLALARGHQFAPFHILKVLBEKXGSLRALSAGR--PD 58

61 SAERVINOALKKLPSQPPDDIPASSSLIKVIRPAAQAQKSRGTHLAVDQLMGLL-E 119

59 QLDGGVETLLAKTPRVGDAGGQLYMKPDTARVFAEAKSAKAAGDAFTVTERLLIAIAKE 118

120 DSQIRDLNVEGVATARVYKSEVEKLRGKGVESASGDTNFQALKTYGRDLVQA--GK 177

119 GGEAAKLFKEAGVSAQSLETAANMR--KGRADSANAEBGYEALKRYARDLTAARDGK 176

178 LDPVIGRDEEIRRVRIISRRTPKNNPVLIGEPGVGTAVVEGLAQIRIVKGDVPNSLTDVR 237

177 LDPVIGRDEEIRRTQVLSRRTPKNNPVLIGEPGVGTAVVEGLALRVNGDVPESLKDCK 236

238 LLSLDMGALVAGAKYRGFEERLKVLEVEDAEKVLFIIDEHLVLGAGKTSGSMDAA 297

237 LLSLDMGSLIAGAKYRGFEERLKVLEVEDAEKVLFIIDEHMTLVGAGKGGAMDAS 296

298 NLFKPLMARGQLRCIGATTLEBYRVYKVDAAFRFQVYVAEVSPTDTSILRGLKEK 357

297 NLLKPALARGEHLGVGATTLBYRKHVSKDAALARPQVFPSEPTVEDTVSILRGLKEK 356

358 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANRVVOLDQSQPEIDNLE 417

357 YEVHGVRISSAIVAAATLSNRYIADRFDPKAIDLVEASSRVQMIDSKPELDEID 416

418 RKMOLIEHLHLEKEKAKARLIEVVKELDDLDLQPLTKMYRKEKERIDEIRLK 477

417 RLVLQKIERALSKEITDAASKQRLLENLEVEIDDLQFRSDENTARWKAEEKVGAQAAR 476

478 OKRELMFSLQEAERYDLARAADLRYGAIQVESAIQLEGTSEENVMLTENVGPHI 537

477 EALDRRLADLANAQAGDFARAGQIQYGEIPALERLAEAE--AGDTQALTPEVVDABQI 534

538 AEVSRWTGIPVTRIGONEKERLIGLADRLKHVVVGQNAVNAVSEALIRSRAGLGRAQ 597

535 AAVSRWTGVPVEKMLEGEREKLLKWEDELGRVVGQDEALEAVSDAVRRARAGLQPSK 594

598 PGSGFLFGPTGVGTAKALAEQIFDENLILVRIDMSYMEQHSRLIGAPPGVGH 657

595 PGSGFLFGPTGVGTAKALAEQIFDENLILVRIDMSYMEQHSRLIGAPPGVGY 654

658 BEGGQLTEAVRRRPVCVILFQVEKAVHVAFTLLQVLDGRLTDGQRTYDFRNSVLIIM 717

655 DEGGALTEAIRRPVQVVLFDIEKADHPDVFNVLLQVLDGRLTDGQRTYDFRNTLIIM 714

718 TSNLGAHLHLAGLTKGVTVNEVAROCVMBREVRKHFPELNLRLDEIWFDPDLSHDLKRA 777

715 TSNLGAELYASQEDGE-DVEAVRPMVMNTVRGHPPEFLNRIIDEIILFKLSRNMGDIV 773

778 RLQKDVAVRLAERGVAVLAVTDAALDYILAESDYQVYCARPIRMWKKVTELKMWVR 837

774 RLQORVEKLADRRMALADAEALNWLADKGYDVPYGARPLKRVIQELVDPIAKKLLA 833

838 EBIDNSNTVYIDAGAGDLVY 857

Db 834 GEIEDGGVIAVGVTGQLSF 853

RESULT 12

AD2441

endopeptidase Clp ATP-binding chain B [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AD2441

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD2441

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-835 <KUR>

A;Cross-references: GB:BA000019; PIDN:BAH76783.1; PID:g17134222; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: clpB

C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.8%; Score 2218.5; DB 2; Length 835;

Best Local Similarity 52.6%; Pred. No. 5.7e-91;

Matches 430; Conservative 181; Mismatches 185; Indels 21; Gaps 7;

Qy 43 TGIFPQAISSAGGENAAQSAERVINOALKKLPSQPPDDIPASSSLIKVIRPAAQAQKS 102

Db 12 SGILTKA-----GVNLQKISDRT-EQYIQRPKVSGNSTSVYLGRSLDILLDRAEAHRKD 65

Qy 103 RGTHLAVDQLINMGLEDSQI-RDILNVEGVATARVYKSEVEKLRGKGVESASGDTNF 161

Db 66 FQDEYISIEHLLLAYPKDDRFKGLFQEFALDESKLNIQVRSQ--TVTDQPEGY 123

Qy 162 QALKTYGRDLVEQA--GKLDPVIGRDEEIRRVVRIISRRTPKNNPVLIGEPGVGTAVVEG 219

Db 124 QSLKCYGRDLTEAARKQGLDPVIGRDEEIRRTIQLISRRTPKNNPVLIGEPGVGTAVVEG 183

Qy 220 LAQRIYKGDVPNSLTDVRLISLDMGALVAGAKYRGFEERLKVLEVEDAEKVLFI 279

Db 184 LAQRIYVAGDVPQSLKDKRLISLDMGAMTAGKFRGEFEERLKVLEVEDAEKVLFI 243

Qy 280 EHLVLGAGKTSGSMDAANLFKPLMARGQLRCIGATTLEBYRVYKVDAAFRFQVY 339

Db 244 EHTVVGAGATQAGMDAGNLLKPLMARGELRCIGATTLEBYRVYKVDAAFRFQVY 303

Qy 340 AEPSVPTTISILRGLKEKYGHHGVRIQDRALINAAQLSARYITGRHLPDKAIDL 399

Db 304 DQPSVEDTTISILRGLKERYEVHGVKISDSLSVAATLSRYSISDRFLPDKAIDL 363

Qy 400 ANRVOLDQSQPEIDNLEKRMOLIEHLHLEKEKAKARLIEVVKELDDLDLQPL 459

Db 364 ARKMEITSKPELDEIDRKILQLEMEKLSQSDAASRELERLEKELADLKEQRTL 423

Qy 460 TMKYRKEKERIDEIRLQKREBELMFLSQEAERYDLARAADLRYG-----AIQVES 512

Db 424 NTQWQSEKQVINKLQSVKEEIDKVNLEIQQAERNVDLNPAAELKYNLTDLHRRLEATER 483

Qy 513 AIAQLGTSSEENVMLTENVGPEHIAEVSVRWGTGIPVTRIGONEKERLIGLADRLKH 572

Db 484 ELSQTQGTGKS---LLREEVTEADIAETISKTGTIPISKLVESEKEKLHLEDELHHRV 540

Qy 573 GQNAQNAVNAVSEALIRSRAGLGRAQQTGSLFGTGVGTAKALAEQIFDENLILVR 632

Db 541 GQDEAVTADVAIQRSRAGLADPNRPTAFVFLGTGVGTAKALAEQIFDENLILVR 600

Qy 633 IDMSYMEQHSRLIGAPPGVGHVEEGQLTEAVRRRPYCVILEDEVEKAVHVAFTNLL 692

Db 601 IDMSYMEKHAVSRILGAPPGVGVVEEGQLTETIRRPYAVILFDEIEKADHPDVNFIL 660

346 KORYEAHGVQITDNAVVAQAQANRYITNRPDPKDAIDLIDBACANVRVQLSSRPEAD 400
 QY 415 NLERKRMQLEIETLHALEREKDKASKARLIEVVRKELDLDRKLOPLTWKRYKKEKERIDEIR 474
 Db 406 ILERKKRQLEETAKALERKEAASRRERKLKVKADIQRVEEELQPLVSKYNDERQRIDELQ 465
 QY 475 RLKQKREBELMFSLQEAERRYDLARAADRLRYGAIQEVESAQAL-EGTSSSENVMLTENVG 533
 Db 466 EQWSRIDBKK-KLERAVRDGQMDLAADLQVNVPLIQDIRISLKEDIERQKATLVOEKVT 524
 QY 534 PPHIAEVVSRWTGIPVTRLQNGEKERLIGLADRLHKRVVGQNOAVNASAILRSRAGLG 593
 Db 525 EGDVAAVAVRWGIPVVKLSQTDRERLNL-SMELHRRVKGQDEAVERVADAIIRARAGLS 584
 QY 594 RAOQTGSFGLFGPTGVGKTELAKALABQLFDENLLVRIIDMSEYMEQHSVSRLLGAPPG 653
 Db 585 RPNSTASFLFGPTGVGKTELKVAVAEELFDDEKHMVRIIDMSEYMEQHSVSRLLGAPPG 644
 QY 654 YVHGEEGQLTEAVRRPYPVCVLLPDEVEKAHVAVENTLLQVLDGRLTDGQGRTVDFRNS 713
 Db 645 YIGHDEGQLTEPVRRPHAVVLFDEVEKAHPVNVVLLQVLDGRLTDSRGRTVDFNS 704
 QY 714 VIIMTSNLGABHLAGITGKVTMEVARDCVMRVRKHPPELNLRLDEIVVDFDPLSHDQL 773
 Db 705 IIVTMSNLGSHLNPBETNESYEVLRNVLAARVSRPELNLRLDDIVVFFRLRTEDL 764
 QY 774 RKVARLOMKDVAVLAERGVALAVTDAALDYVLAESYDVPYVGARPIRRMWEKVVTELK 833
 Db 765 RGVDNLTAGVNERKLSGFSVLLDGVKDFILHGHGDANMGARPLRRWIEKNIVTEIGR 824
 QY 834 MVRREIDENST--VYIDAGAGDLVYRVESG 862
 Db 825 MLIAKELPPNSTLRVSLPEGGNKLTGCVKRG 855

RESULT 14
 AI0831
 ClpB protein (heat shock protein f84.1) [imported] - Salmonella enterica subsp.
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #ext_change 18-Nov-2002
 C:Accession: AJ0831
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.;
 th, T.; Connerston, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.
 , S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stee
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enter
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AI0831
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-857 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD05840.1; PID:gi650381.5; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY2849
 C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.4%; Score 2204; DB 2; Length 857;
 Best Local Similarity 52.1%; Pred. No. 2.6e-90;
 Matches 449; Conservative 155; Mismatches 244; Indels 14; Gaps 9;

QY 1 MNPKEFKTKNETIATAHELAVNAGHAQPTPLHLAGALISDPTGIFPQAISSAGGENAAQ 60
 Db 1 MRDLRNLKPOLADLAQSLAHGDNQFLEPLHMSALLNQGSGSIRPLTSA-GINAGQ 59
 QY 61 SAERVINQALKKPLPSQSPPPDDIPASSLLIKYVRAAQAQSRGDTHLAVDQLTMGLLED 120
 Db 60 -LRTAIDQALSRLPQVEGTGGDQVPSSELVRVNLCDLAKRGDGNFISSELFVAIALES 118
 QY 121 -SQIRDLLNVEGVATARVKSVEKFLGKGGKSVESAGDTNFQALTKYTGRLDVEQA--GK 177
 Db 119 RGTLDLLKAGATTANITQALEONRG--GESVNDQGAEDORQALKKYTVLTPTRASQK 176

178 LDPVIGDEEIRRVIRLRSRTKNNPVLIGEPGVGKTA VVEGLAQRIKGVDPNSLTDV 237
 177 LDPVIGDEEIRRVIRLRSRTKNNPVLIGEPGVGKTA VVEGLAQRIKGVDPNSLTDV 236
 238 LISIDMGALVAGAKYRGEFEERLKSVLKEVEDAEKVVILFIDEIHLVLAGKTEGSDMA 297
 237 VLLIDMGALVAGAKYRGEFEERLKSVLKEVEDAEKVVILFIDEIHLVLAGKTEGSDMA 296
 298 NLKPKMLARQLRCIGATTLEERKVKVEKDAEAFERRFOQVYVAEPSPVDTISILRLKE 357
 297 NMLKPALARGELHCVGATTLDEVRKVEKDAEAFERRFOQVYVAEPSPVDTISILRLKE 356
 358 YEGHGVRIODRALINAAQLSARYITGRHLDPKADLVDEACANVRVQLDSOPEEIDNL 417
 357 YELHHVQITDPAIVAAATLSHRYIADRLPDKADLVDEACANVRVQLDSOPEEIDNL 416
 418 RKMOLIEIHLALEREKDKASKARLLIEVRKELDLDRKLPKTMKYEKERIDEIRLKE 477
 417 RRIIQLKLEQALMKSEDEASKRRLDMLNEELDDKERQYSELEEWKAEKASLSGTQTIK 476
 478 QKBEELMFSIQEAREYRLARADLRYGAIQEVESAIQALE-GTSSEENVN--LTENVGP 534
 477 AELEQAKIAIEQARRVGDLMSELOVGIPELEK--QLEAATQSEKTMRLLRNKVTD 533
 535 EHTAEVSVRWGTGIPVTRLGQNEKERLIGLADRLHKRVGQONQAVNAISAILRSRAGLR 594
 534 AETAELVARTGIPVSRMLEGERELRMEQELHSRVIGQNEAVEAVSNARRSAGLSD 593
 595 AQPTGSFLFLGPTGVGKTELAKALAEOLFDENLVRIDMSYMEHOSVRLIGAPPCY 654
 594 PNEPISGFLFLGPTGVGKTELAKALAEOLFDENLVRIDMSYMEHOSVRLIGAPPCY 653
 655 VGEHGGQTEAVRRPYPYCVILFDEVEKHAHVAVNTLLQVLDGRLTDGQRTVDPRNSV 714
 654 VGYEEGYLTEAVRRPYPYCVILFDEVEKHAHVAVNTLLQVLDGRLTDGQRTVDPRNSV 713
 715 IIMTSLNGAHLHLAGITGKVTMEVARDVCMVRKVEKHAHVAVNTLLQVLDGRLTDGQRTVDPRNSV 774
 714 VIMTSLNGSD-LIQERFGLDYGKMEKVLGVVYQNSQNFPEFNRIDEVVFHPLGEQHIA 772
 775 KVARLQMKDVAVRLAERGVAVATDAALDYLAESYDVPVYGPAPIRRMWKKVVTLSK 834
 773 SIAQIQOLQRYKLEBERGYEIHISDALKLLSANGYDVPYGPAPIRRMWKKVVTLSK 832
 835 VREIDENSTVIDAGAGLV 856
 833 ILSGELVPGKVRLEANDRIV 854

SULT 15
 3070

Species: Agrobacterium tumefaciens
 Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 Accession: AC3070

Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 Gege, G.; Giller, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Karp, P.; Romero, P.; Zhang, S.
 ence 294, 2317-2323, 2001
 Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 er, E.W.

Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 Reference number: AB2577; MUID:21608550; PMID:11743193
 Accession: AC3070
 Status: preliminary
 Molecule type: DNA
 Residues: 1-874 <KUR>
 Cross-references: GB:AE008689; PID:AA14977.1; PID:g17742634; GSPDB:GN00187
 Experimental source: strain C58 (Dupont)
 Genetics:
 Gene: clpB
 Map position: linear chromosome

C: Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.3%; Score 2199; DB 2; Length 874;
 Best Local Similarity 51.7%; Pred. No. 4.e-90;
 Matches 447; Conservative 155; Mismatches 253; Indels 10; Gaps 6;

QY 1 MNPKFTHTKNTTATTAHELAVNAGHAQFTPLHLAGALISDPTGIFFPAISSAGGENAAQ 60
 DB 1 MNIDKYSERVGFQSAQTFALAEHQFSEHVLKVLDDQEQMAASLIERAGD--AK 58
 QY 61 SAERVINOALKKLPSSQPPDDIPASSSLIKVIRRAQAOKSGRDTHLAVDQLIMGLLED 120
 DB 59 EARLANDAALAKLPKVGSGNGLSLTAPLAKVFSTARDLAKAGDSFVTVVERLLQALAE 118
 QY 121 SQIRD--LLNEVGAVATARVKEVEKLGKGEKGVESAGDTNFQALATYGRDIVEQA--G 176
 DB 119 SSASTSASLKAGATAQALNQVINDIR--KGTADSANAQGFQALAKYARDLITEAREG 176
 QY 177 KLDPVIGDEEIRRVIRLRSRTKNNPVLIGEPGVGKTA VVEGLAQRIKGVDPNSLTDV 236
 DB 177 RLDPVIGRDEIRRTIQVLSRRTKNNPVLIGEPGVGKTA VVEGLAQRIKGVDPNSLTDV 236
 QY 237 RLISIDMGALVAGAKYRGEFEERLKSVLKEVEDAEKVVILFIDEIHLVLAGKTEGSDMA 296
 DB 237 KLMDLMGALVAGAKYRGEFEERLKSVLKEVEDAEKVVILFIDEIHLVLAGKTEGSDMA 296
 QY 297 ANLFPKMLARQLRCIGATTLEERKVKVEKDAEAFERRFOQVYVAEPSPVDTISILRLKE 356
 DB 297 SNLLKPALARGELHCVGATTLDEVRKVEKDAEAFERRFOQVYVAEPSPVDTISILRLKE 356
 QY 357 YEGHGVRIODRALINAAQLSARYITGRHLDPKADLVDEACANVRVQLDSOPEEIDNL 416
 DB 357 KYEGRHKVIRISDSALVAAATLSNRYITDRFLPKADLVDEACANVRVQLDSOPEEIDNL 416
 QY 417 ERKMOLEIHLALEREKDKASKARLLIEVRKELDLDRKLPKTMKYEKERIDEIRL 476
 DB 417 DRIIQLKIEREALKQETDQSSVDRLKLEDELADTEBKADALTARMAEKQKLGHADL 476
 QY 477 KQBEELMFSIQEAREYRLARADLRYGAIQEVESAIQALEG--TSSEENVNLTENVGPE 535
 DB 477 KKRLEARNELAIQRNGQFQAGELTYGII PGLKELAAAEARDSSGAGSMVQEVVTPD 536
 QY 536 HIAVSVRWGTGIPVTRLGQNEKERLIGLADRLHKRVGQONQAVNAISAILRSRAGLR 595
 DB 537 NIAHVSRWGTGIPVTRLGQNEKERLIGLADRLHKRVGQONQAVNAISAILRSRAGLR 596
 QY 596 QOPTGSFLFLGPTGVGKTELAKALAEOLFDENLVRIDMSYMEHOSVRLIGAPPCY 655
 DB 597 NRPIGSFLFLGPTGVGKTELAKALAEOLFDENLVRIDMSYMEHOSVRLIGAPPCY 656
 QY 656 GHEGGQTEAVRRPYPYCVILFDEVEKHAHVAVNTLLQVLDGRLTDGQRTVDPRNSV 715
 DB 657 GYEGGQTEAVRRPYPYCVILFDEVEKHAHVAVNTLLQVLDGRLTDGQRTVDPRNSV 716
 QY 716 IIMTSLNGAHLHLAGITGKVTMEVARDVCMVRKVEKHAHVAVNTLLQVLDGRLTDGQRTVDPRNSV 775
 DB 717 IIMTSLNGSD-FMTOMGNDNDVDSVRELVMERVSRHPEFNRIDEVVFHPLRDRDEMA 775
 QY 776 VARLQMKDVAVRLAERGVAVATDAALDYLAESYDVPVYGPAPIRRMWKKVVTLSK 835
 DB 776 IVEIQLKRLVSLADRLKLEDEDEARSWLANKGYDPYGPAPIRRMWKKVVTLSK 835
 QY 836 VREIDENSTVIDAGAGLVYRVE 860
 DB 836 LGGEIPDGSRVKVTSGTDRLLFKVK 860

Search completed: February 13, 2004, 01:34:13
 Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 13, 2004, 00:52:50 ; Search time 39 Seconds
(without alignments)
1098.496 Million cell updates/sec

le: US-09-812-350-17
fect score: 4550
uence: 1 MNPKEFKTKNETIATAHEL.....VKMKRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

urched: 127863 seqs, 47026705 residues

al number of hits satisfying chosen parameters: 127863

imum DB seq length: 0

imum DB seq length: 2000000000

rocessing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	Length	DB ID	Description
1	4545	99.9	911	1 H101 ARATH	P42730 arabidopsis
2	2288	50.3	872	1 CLPB SYN3	P74361 synchocyst
3	2254	48.9	868	1 CLP TRVB	P31543 trypanosoma
4	2224	48.9	883	1 CLPB SYN7	P53533 synchococc
5	2215.5	48.7	854	1 CLPB THETH	P9ra63 thermus the
6	2193	48.2	857	1 CLPB ECOLI	P03815 escherichia
7	2162.5	47.5	856	1 CLPB HASIN	P44403 haemophilus
8	2158.5	47.4	848	1 CLPB MYCTU	O53719 mycobacteri
9	2145	47.1	852	1 CLPB CORGL	P53532 corynebacte
10	2104	46.2	878	1 CLPB TRSPA	O83110 treponema p
11	2039	44.8	860	1 CLPB BACNO	P17422 bacteroides
12	1890	41.5	908	1 H104 YEAST	P31539 saccharomyc
13	1884.5	41.4	856	1 CLPB HELPJ	O9zmh1 helicobacte
14	1876.5	41.2	856	1 CLPB HELPY	P71404 helicobacte
15	1844.5	40.5	812	1 HS78 CANAL	O96ux5 candida alb
16	1805.5	39.7	811	1 HSP7 YEAST	P33416 saccharomyc
17	1759.5	38.7	810	1 CLPB BACSU	P37571 bacillus su
18	1759.5	38.7	848	1 CLPB MYCTU	O06286 mycobacteri
19	1752.5	38.5	848	1 CLPB MYCLE	P24428 mycobacteri
20	1737.5	38.2	819	1 CLPC GULTH	O78410 guillardia
21	1735.5	38.1	922	1 CLPA PEA	P35100 pisum sativ
22	1733.5	38.1	854	1 CLPC CYACA	O9tm05 cyanidium c
23	1731	38.0	923	1 CLAB LYCES	P31542 lycopersico
24	1703.5	37.4	926	1 CLAA LYCES	P51341 lycopersico
25	1697.5	37.3	821	1 CLPC PORPU	P51332 porphyra pu
26	1658	36.4	714	1 CLPB MYCGE	P75297 mycoplasma
27	1655	36.4	715	1 CLPB MYCPN	P46523 mycoplasma
28	1654	36.4	874	1 CLPA BRANA	O54316 treponema h
29	1639.5	36.0	828	1 HLYB TREHY	O9pkas chlamydia m
30	1613.5	35.5	870	1 CLPC CHLMU	O9z8a6 chlamydia h
31	1602	35.2	845	1 CLPC CHLUPN	O84288 chlamydia p
32	1602	35.2	854	1 CLPC CHLTR	P49574 odontella s
33	1547.5	34.0	885	1 CLPC ODOSI	

ALIGNMENTS

RESULT 1

ID	H101 ARATH	STANDARD;	PRT;	911 AA.
AC	P42730; Q8W4F2; Q9LE57;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Heat shock protein 101.			
GN	HSP101 OR AT1G74310 OR F1017.2.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia; TISSUE=Leaf;			
RX	MEDLINE=95170291; PubMed=7866032;			
RA	Schirmer E.C., Lindquist S., Vierling E.;			
RT	"An Arabidopsis heat shock protein complements a thermotolerance defect in yeast."			
RL	Plant Cell 6:1899-1909(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=20226114; PubMed=10760305;			
RA	Hong S.-W., Vierling E.;			
RT	"Mutants of Arabidopsis thaliana defective in the acquisition of tolerance to high temperature stress."			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4392-4397(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=21016719; PubMed=11130712;			
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A., Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.B., Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;			
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana."			
RL	Nature 408:816-820(2000).			
RN	[4]			
RP	SEQUENCE OF 1-460 FROM N.A.			
RC	STRAIN=cv. Columbia;			

P42762 arabidopsis
Q9S522 lactococcus
Q9C109 lactococcus
P35594 streptococ
P15716 escherichia
P05444 rhodopseudo
Q06716 lactococcus
Q0649 mus muscullu
Q9h078 homo sapien
Q9wt2 rattus norv
Q01357 rhodococcus
Q51416 pseudomonas

34 1478.5 32.5 945 1 ERD1 ARATH
35 1422 31.3 748 1 CLPE LACLC
36 1418 31.2 748 1 CLPE LACLC
37 1396.5 30.7 752 1 CLPE STRPN
38 1331.5 29.3 758 1 CLPA ECOLI
39 1299.5 28.6 793 1 CLPA RHOBL
40 1256 27.6 763 1 CLPL LACLA
41 534.5 11.7 677 1 SKD3 MOUSE
42 534.5 11.7 707 1 SKD3 HUMAN
43 525.5 11.5 677 1 SKD3 RAT
44 463.5 10.2 351 1 ADPR_RHOER
45 421 9.3 371 1 AMIB_PSEAE

Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 SSP consortium (Salk/Stanford/PCEC)";
 Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THERMOTOLERANCE.
 -!- INDUCTION: By heat shock.
 -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U13949; AAA67927.1; --
 EMBL; AF218796; AAF26423.1; --
 EMBL; AC020579; AAG52410.1; --
 EMBL; AY062596; AAL32674.1; --
 PIR; F96771; F96771.
 InterPro; IPR003593; AAA ATPase.
 InterPro; IPR001270; Chaprin-clpa/b.
 InterPro; IPR004176; Clp_N.
 Pfam; PF00004; AAA; 2.
 Pfam; PF02861; Clp_N; 2.
 SMART; SM00382; AAA; 2.
 SMART; SM00382; AAA; 2.
 PROSITE; PS00870; CLPB_1; 1.
 PROSITE; PS00871; CLPB_2; 1.
 Chapterone; Heat shock; ATP-binding; Repeat.
 DOMAIN 164 410
 NP_BIND 532 723
 NP_BIND 207 214
 NP_BIND 606 613
 NP_BIND 141 141
 CONFLICT 595 595
 CONFLICT 595 595
 P -> F (IN REF. 4).
 P -> A (IN REF. 1).
 SEQUENCE 911 AA; 101294 MW; 191EC1853B0C4CB9 CRC64;
 Query Match 99.9%; Score 4545; DB 1; Length 911;
 Best Local Similarity 99.9%; Pred.No. 3.2e-191;
 Matches 910; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 1 MNPEFTHKNTETIATAHELAVNAGHAQTPHLAGALISDPTGIFPQAISAGGENAAQ 60
 1 MNPEFTHKNTETIATAHELAVNAGHAQTPHLAGALISDPTGIFPQAISAGGENAAQ 60
 61 SAERVINQALKLPQSPPDDIPASSSLIKVIRRAQAQKSRGDTHTLAVDQLIMGLLED 120
 61 SAERVINQALKLPQSPPDDIPASSSLIKVIRRAQAQKSRGDTHTLAVDQLIMGLLED 120
 121 SQIRDLLENGVATARKVSEKLGKGGKVESAGDTNFQALKTYGRDLVEQAGKLDLP 180
 121 SQIRDLLENGVATARKVSEKLGKGGKVESAGDTNFQALKTYGRDLVEQAGKLDLP 180
 181 VTGRDEIRRVRLISRRTPKNNPVLIGEPGVKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
 181 VTGRDEIRRVRLISRRTPKNNPVLIGEPGVKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
 241 LDMGALVAGAKYRGEFEELKSVLKEVEDAEQKVFIDEHLVLGAGTKGSDMAANLF 300
 241 LDMGALVAGAKYRGEFEELKSVLKEVEDAEQKVFIDEHLVLGAGTKGSDMAANLF 300
 301 KPMALRGQLRCIGATTLEBYRYKVEKDAFERFQVYVAEPSVDTISILRGLKEKVEG 360
 301 KPMALRGQLRCIGATTLEBYRYKVEKDAFERFQVYVAEPSVDTISILRGLKEKVEG 360
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDCAANVRVQLDSQPEIDNLERK 420
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDCAANVRVQLDSQPEIDNLERK 420
 421 MOLETELHALEREKOKAKARLIEVRKELDDLDKLOPLTMKYRKEKERIDEIRLQKQR 480

Db 421 MOLETELHALEREKOKAKARLIEVRKELDDLDKLOPLTMKYRKEKERIDEIRLQKQR 480
 Qy 481 EELMFSLOEAEERRYDLARAADLYGALQEVESAIAQLEGTSSENVMLTENVGPEHIAEV 540
 Db 481 EELMFSLOEAEERRYDLARAADLYGALQEVESAIAQLEGTSSENVMLTENVGPEHIAEV 540
 Qy 541 VSRWTGIPVTRLGQNEKERLIGLADRLHKEVGVQGNQAVNAVSEAILRSRAGLRGPOPTG 600
 Db 541 VSRWTGIPVTRLGQNEKERLIGLADRLHKEVGVQGNQAVNAVSEAILRSRAGLRGPOPTG 600
 Qy 601 SFLPLGTGVTGKTELAKALAEQFDDENLIVRIMGEYMEQHSVRLIGAPPYVGHHEG 660
 Db 601 SFLPLGTGVTGKTELAKALAEQFDDENLIVRIMGEYMEQHSVRLIGAPPYVGHHEG 660
 Qy 661 GQLTEAVRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGQGTVDPRNSVIITSN 720
 Db 661 GQLTEAVRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGQGTVDPRNSVIITSN 720
 Qy 721 LGAEHLLAGLTGKVTMEVARDCVMRVVRKHFRPELNLRLDEIVVFDPLSHDQLRKVARLQ 780
 Db 721 LGAEHLLAGLTGKVTMEVARDCVMRVVRKHFRPELNLRLDEIVVFDPLSHDQLRKVARLQ 780
 Qy 781 MKOVAVRLAERGVALAVTDAAALYILAESYDVPYVGARPIRRMMEKKVVTLSKVVREI 840
 Db 781 MKOVAVRLAERGVALAVTDAAALYILAESYDVPYVGARPIRRMMEKKVVTLSKVVREI 840
 Qy 841 DENSTVYIDAGDLVYRVESGGLVDASTGKSDVLIHTANGPKRSDAAQAKVKMEIEI 900
 Db 841 DENSTVYIDAGDLVYRVESGGLVDASTGKSDVLIHTANGPKRSDAAQAKVKMEIEI 900
 Qy 901 EDDNEEMIED 911
 Db 901 EDDNEEMIED 911
 RESULT 2
 CLPB SYN3
 ID CLPB SYN3 STANDARD; PRT; 872 AA.
 AC P74361;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ClpB protein.
 GN CLPB OR SLR1641.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RK MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima T., Hiroseawa M., Sugiyura M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
 CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AB012390; BAA81745 1: --

706 LRSNLGSPLEGLQKGMPEYRDEIRVFKVLOQHPFELNRLDRIIVFPLTKRQIRQI 765
777 ARLOKQVAVRAERGVAVLADVAALDYLAESYDPVVGARPIREWEKKVVTLSKVV 836
766 VEQISYLRARLAERKRISLLELTAERQFLAERGYPDPVVGARPIRVIORELETPAQKIL 825
837 REEIDENSTVYIDAGAGDVLVYRV 859
826 AGEVKEGDRVQVDVGPAGLVFVAV 848

SEQUENCE FROM N.A.
MEDLINE=90239044; PubMed=2185473;
Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M.,
Mattick J.S., Dalrymple B., Kuramitsu H., Shiroza T., Foster T.,
Clark W.P., Ross B., Squires C.L., Maurizi M.R.;
"Conservation of the regulatory subunit for the Clp ATP-dependent
protease in prokaryotes and eukaryotes";
Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517 (1990).

[1]
SEQUENCE FROM N.A.
MEDLINE=90239044; PubMed=2185473;
Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M.,
Mattick J.S., Dalrymple B., Kuramitsu H., Shiroza T., Foster T.,
Clark W.P., Ross B., Squires C.L., Maurizi M.R.;
"Conservation of the regulatory subunit for the Clp ATP-dependent
protease in prokaryotes and eukaryotes";
Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517 (1990).

[2]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12";
Science 277:1453-1474 (1997).

[3]
SEQUENCE FROM N.A.
STRAIN=K12;
MEDLINE=97349980; PubMed=9205837;
Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubraman S.,
Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
Yamagata S., Horiuchi T.;
"Construction of a contiguous 874-kb sequence of the Escherichia coli
K-12 genome corresponding to 50.0-68.8 min on the linkage map and
analysis of its sequence features";
DNA Res. 4:91-113 (1997).

[4]
SEQUENCE FROM N.A.
STRAIN=O6:H1 / CFT073 / ATCC 700928;
MEDLINE=22388234; PubMed=12471157;
Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
"Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli";
Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

[5]
SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Pofst G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grofbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533 (2001).

[6]
SEQUENCE FROM N.A.
STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22 (2001).

[7]
SEQUENCE OF 1-593 FROM N.A.
STRAIN=K12;
RX MEDLINE=91294165; PubMed=1906060;
Kitagawa M., Wada C., Yoshioka S., Yura T.;
RT "Expression of ClpB, an analog of the ATP-dependent protease
regulatory subunit in Escherichia coli, is controlled by a heat shock
sigma factor (sigma 32)";
EL J. Bacteriol. 173:4247-4253 (1991).

[8]
SEQUENCE OF 753-857 FROM N.A.
MEDLINE=82247208; PubMed=6285294;
Shen W.-F., Squires C., Squires C.L.;
RT "Nucleotide sequence of the rrmG ribosomal RNA promoter region of
Escherichia coli";
RL Nucleic Acids Res. 10:3303-3313 (1982).

[9]
SEQUENCE OF 1-31 FROM N.A.
STRAIN=K12 / W3110;
RA Ogura T., Tomoyasu T.;
EL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

[10]
SEQUENCE OF 1-14; 150-157; 355-364 AND 452-460.
MEDLINE=92062147; PubMed=1953774;
RA Pontis E., Sun X.Y., Joernvall H., Krook M., Reichard P.;
RT "ClpB proteins copurify with the anaerobic Escherichia coli
reductase";
EL Biochem. Biophys. Res. Commun. 180:1222-1226 (1991).

[11]
IDENTIFICATION AS A HEAT SHOCK PROTEIN.
MEDLINE=91294166; PubMed=2066329;
RA Squires C.L., Pedersen S., Ross B.M., Squires C.;
RT "ClpB is the Escherichia coli heat shock protein F84.1";
EL J. Bacteriol. 173:4254-4262 (1991).

CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR ATP-
DEPENDENT PROTEASE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, ClpB (shown here) and ClpB-3, are produced
CC by alternative initiation;
CC -1- INDUCTION: By heat shock.
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; M29364; AAA24422.1; -

[illegible]

PIR; F64098; F64098.
TIGR; H10859; --
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_central.
InterPro; IPR001270; Chaprinin_clpA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 2.
PRINTS; PRO0300; CLPPTOTASEA.
SMART; SMO0382; AAA; 2.
PROSITE; PS00870; CLPAB_1; 1.
PROSITE; PS00871; CLPAB_2; 1.
Chaperone; ATP-binding; Repeat; Complete proteome.
DOMAIN 161 409 I.
DOMAIN 531 722 II.
NP_BIND 206 213 ATP (POTENTIAL).
NP_BIND 605 612 ATP (POTENTIAL).
SEQUENCE 856 AA; 95837 MW; B8958ED9BD03EA3B CRC64;

Query Match 47.5%; Score 2162.5; DB 1; Length 856;
Best Local Similarity 52.1%; Pred. No. 2e-87;
atches 444; Conservative 153; Mismatches 245; Indels 11; Gaps 8;

1 MNPEKFTHTKTETATAHELAVNAGHAQFTPLHLAGALISDPTG-IPPOAIISSAGGENAA 59
1 MNIEKFTTKFOEALSERQSLGKNDQFIEPVHLLTALLNQGGSIAP--ILTAGGVNVA 58
60 QSAERVINQALKLPSSPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILMLLE 119
59 -LLRNELKTELNKLPQVINGDVQLGRQLNLLNLCCKFAQQNQDKFISSEFLFAALE 117
120 D-SQIRDLNNEGVVATARKVESEVKEKLKGGKGVESAGDTNFOALTYGRDIVEQA--G 176
118 ERGTISDILKCGAKKQEQISQAIQHIG--QGVNDQNAEESRQALEKYITDILTARAESG 175
177 KLDVIGRDEIRVRLSRRTKNNPVLIGEPGVKTAVVEGLAQRIKVGDPNSLTDV 236
176 KLDVIGRDEIRAIQVLRQTKNNPVLIGEPGVKTAVVEGLAQRIKVGDPNSLTDV 235
237 RLISLDGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGTSGMDA 296
236 RVLSLDGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGTSGMDA 295
297 ANLKPMLARQLKICGATTLERVKYVEKDAERFRQVYVAEPVPTISILRLKE 356
296 GNLLKPSLARGELCHVGATTLDEYRQVIEKDAERFRQVYVAEPVPTISILRLKE 355
357 KYEGHGVIRIDRALINAAQLSARYITGRHLPDKAIDLVDEACANVRVQDSPEIDNL 416
356 RYEHVHVDITDPAIVAAATLSHRYISDRQPLKAIIDLEAASSIWEIDSKPELDRL 415
417 ERKMQLEIEUHALERKOKASKARLIEVRKELDDRLKPLQPLTKYRKEKERIDIRRL 476
416 ERTIQLKLEQALQKQBEDEASRLKMLEKAEKREVAEEVWVSEKATLSGSHI 475
477 KQREELMFSIQEARRVLDLARAADLYGALQEVESAIAOLEGTSSEENVMLTENGPEH 536
476 KQELDTAKTELEARRAGDLANSELQIGRPDLLEKQLEQAEYSEGKEMTLRYRYTDEE 535
537 IAEVSVRWITGPIVTRLQNGEKERLIGLADRLHKRVVQGNQAVNAISAILRSRAGLGRAQ 596
536 IAEVLSKATGIPVSKMMEGEKEKLLRMEDELHKEVIGQEEAVDAVANAIRSRAGLSDPN 595
597 QPTGSFLPGTGVGKTELAALAEQFLDDENLLVRIDMSEYMEQHSVSLIAGPGYVG 656
596 RPIGSFLPGTGVGKTELAALAEQFLDDENLLVRIDMSEYMEQHSVSLIAGPGYVG 655
657 HEEGQLEAVERRPYCVILPDEVEKAHVAVENTLLOVLDGRLTDQGRGTVDPRNSVII 716
656 YEEGYLEAVERRPYCVILPDEVEKAHVAVENTLLOVLDGRLTDQGRGTVDPRNSVII 715
717 MTSNLGAELHLAGITGKVTMEVARDCVMRKFRPELLNRLDEIVVFDPLSHDQLRKV 776

PIR; F64098; F64098.
TIGR; H10859; --
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_Centr.
InterPro; IPR001270; Chaprtnin_clpA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 2.
PRINTS; PRO0300; CLPPTOTASEA.
SMART; SMO0382; AAA; 2.
PROSITE; PS00870; CLPAB_1; 1.
PROSITE; PS00871; CLPAB_2; 1.
Chaperone; ATP-binding; Repeat; Complete proteome.
DOMAIN 161 409 I.
DOMAIN 531 722 II.
NP_BIND 206 213 ATP (POTENTIAL).
NP_BIND 605 612 ATP (POTENTIAL).
SEQUENCE 856 AA; 95837 MW; B8958ED9BD03EA3B CRC64;

Query Match 47.5%; Score 2162.5; DB 1; Length 856;
est Local Similarity 52.1%; Pred. No. 2e-87;
atches 444; Conservative 153; Mismatches 245; Indels 11; Gaps 8;

1 MNPEKFTHTTETAHAEVLAVNAGHAQFTPLHLAGALISDPTG-IPPOAIISSAGGENAA 59
1 MNIEKFTTKFOEALSERQSLGKNDQFIEPVHLLTALLNQGGSIAP--ILTAGGVNVA 58
60 QSAERVINQALKLPSSPPDDIPASSLLIKVIRRAQAQKSGRDTHLAVDQILMLLE 119
59 -LLRNELKTELNKLPQVINGGVQLGRQLNLLNLCCKFAQQNQDKFISSEFLFAALE 117
120 D-SQIRDLNNEGVVATARKVESEVEKLKGGKGVESAGDTNFQALTYGRDIVEQA--G 176
118 ERGTISDILKCGAKKQEQISQAIQHIG--QGVNDQNAEESRQALEKYITDILTARAESG 175
177 KLDVIGRDEIRVRLSRRTKNNPVLIGEPGVKTAVVEGLAQRIKVGDPNSLTDV 236
176 KLDVIGRDEIRAIQVLRQTKNNPVLIGEPGVKTAVVEGLAQRIKVGDPNSLTDV 235
237 RLISLDGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGTEGSMDA 296
236 RVLSLDGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGTEGSMDA 295
297 ANLKPMLARQLRCIGATTLERVKYVEKDAERFRFQVYVAEVPVPTISILRLKE 356
296 GNLLKPSLARGELCHVGATTLDEYRQVIEKDAERFRFQVYVAEVPVPTISILRLKE 355
357 KYEGHGVIRIDRALINAAQLSARYITGRHLPDKAIDLVDEACANVRVQSDSPEIDNL 416
356 RYEHVHHVDITDPAIVAAATLSHRYISDRQPLKAIIDLEAASSIWEIDSKPELDRL 415
417 ERKMQLEIEUHALERKOKASKARLIEVRKELDDRLKPLQPLTKYRKEKERIDIRRL 476
416 ERTIQLKLEQALQKQBEDEASRLRMLKLEAEKREVAEEVWVSEKATLSGSHI 475
477 KQREELMFSIQEARRVLDLARAADLYGALQEVESAIAOLEGTSSEENVMLTENVGPEH 536
476 KQELDTAKTELEARRAGDLANMSELQTRIPDLLEKQLEQAEYSEGKEMTLRYRYTDEE 535
537 IAEVSVRWITGPIVTRLQNGEKERLIGLADRLHKRVVGQNOAVNAISAILRSRAGLGRAQ 596
536 IAEVLSKATGIPVSKMMEGEKEKLLRMEDELHKEVIGQEEAVDAVANAIRSRAGLSDPN 595
597 QPTGSFLFGTGVGKTELAALAEQFLDDENLLVRIDMSEYMEQHSVSLIAGPGYVG 656
596 RPIGSFLFGTGVGKTECLKTAKFLFDSEDAMVRIDMSEYMEQHSVSLVAGPGYVG 655
657 HEEGQQLTEAVRRPYPYCVILPDEVEKAHVAVENTLLQVLDGRLTDQGRGTVDPRNSVII 716
656 YEEGGLTEAVRRPYPYCVILPDEVEKAHVAVENTLLQVLDGRLTDQGRGTVDPRNTVII 715
717 MTSNLGAELHLAGITGKVTMEVARDCVMRVYRKHFRPELLNRLDEIVVFDPLSHDQLRKV 776

[illegible]

117 LLE-DSQIRDLNVEGVATARKVSEKLRGKGGKVESASGDTNFQALKTYGRDLVEQA 175
 116 IARGKSDAADLTKNGATYDAIKEAPSRVGSQ--RVTTQDPGQFQALEKYSTDLTKLA 173
 176 --CKLDPVLGRDEIRRVRLISRTKNNPVLIGBPGVKTAVVGLAORIVKGVPSNL 233
 174 REGKIDPVLGRDEIRRVVQVLSRRTKNNPVLIGBPGVKTAVVGLARRIVAGDVPSL 233
 234 TDVRLSLDMGALVAGAKYRGFEERLKSVLKEVEDAEAGKVLFDIHLVLCAGKT--EG 292
 234 KGKTLISLDLSGWSVAGAKYRGFEERLKAVIDEIKGANGEVTFIDELHTIVGASGES 293
 293 SDAANLFPMLARGQLRGICGATTLLEYKYVKEKDAAFRRRQQVYVAPSPVDTISLR 352
 294 AMDAGNMKPLLARGELRLVGATTLNEYKYTEKDAALERRRQQVYVGEPTVEDAIGLR 353
 353 GLKEKVEGHGVRIODRALINAQLSARVITGRHLDPKALDVIDEACANVRVOLDSPBE 412
 354 GLKERVEVHGVRIOQSALVAELSNRIITSRFLPDKAIDVDDEASRLRMEIDSSPOE 413
 413 IDNLRKRMQLEIHLALEREKDKASKARLIEVRKELDDRLDKLOPLTMVKYKKEKIDE 472
 414 IDELERIVRLEIEEMALSKESDAASKERLEKRLSELADEREKLSLKARWQNEKTAIDD 473
 473 IRLKQKRELMFSLQEAARRVDLARAADRLRYCAIQEVSAITAQLEG--TSSEENVMLTEN 531
 474 VREMKEELEALSSESIAERDNGYGRVARELYGRIPLEKQTEDAESKVEVVENAMLTSE 533
 532 VQPEHTAEVVSRTGTPVTRLGONKEREKLGILADRLKRVVQGNQAVNAVSEAILRSAG 591
 534 VTPDTIADVVSRTGIPAGKMMQGETKLLNMERVLGNRVQGLEAVTAVSDAVRSRAG 593
 592 LGRAOQPTGSFLGPTGVGKTELAKALAEQFDDENLLVRLDMSEYMEQHSRSLIGAP 651
 594 VADPNRPTGSFLGPTGVGKTELAKAVAEFLFDDDDRAMIRIDMSEYGEKHSVARLVGAP 653
 652 PGVYVHEEGQLTEAVRRPVCVILFDEVEKHAFAVNTLLOVLDGRLTGQGRVDFR 711
 654 PGVYVYDQGGQLTEAVRRPVTVLVDEVEKAHPDVFILLQVLDEGRLLTGQGRVDFR 713
 712 NSVIIMTSLNLAHLLAGLTGKVTMEVARDCVMRVVRKHFRPELLNRLDEIVFPDLSHD 771
 714 NTILLTSLNG-----AGT-----REQMDAVKAFKPEFVNRLLDDVVDFRLSPE 760
 772 QLRKVARLQMKDVAVRLAERGVAVLAVTDAALDYILAESYDPYVGARPIRRWMEKXVTEL 831
 761 QLTISVIDIQIKQLDRLAGRLNRLNRSVDSAKAWLAERGVDPAYGARPLRLIQQAGITL 820
 832 SKWVYREEIDENSTYVIDAGADLVYRVESGGLVDASTGKXSDV 875
 821 AKELLAGNRGQGVLDVVDAG-----GQKLDV 848

ULT 10

B TREPA

CLPB TREPA

STANDARD;

PRT; 878 AA.

083110;

30-MAY-2000 (Rel. 39, Created)

16-OCT-2001 (Rel. 40, Last sequence update)

ClpB protein.

CLPB OR TP0071.

Treponema pallidum.

Bacteria; Spirochaetes;

NCBI TaxID=160;

[1]

SEQUENCE FROM N.A.

STRAIN=Nichols;

MEDLINE=98332770; PubMed=9665876;

Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,

Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RL spirochete";
 RL Science 281:375-388(1998).
 CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
 CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE001191; AAC65062.1; -.
 DR PIR; G71371; G71371.
 DR TIGR; TP0071; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003955; AAA_ATPase_Centr.
 DR InterPro; IPR001270; Chaperin_ClpA/B.
 DR InterPro; IPR004176; Clp_N.
 DR Pfam; PF00004; AAA; 1.
 DR Pfam; PF02861; Clp_N; 2.
 DR PRINTS; PR00300; CLP_PROTEASRA.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00870; CLPAB_1; 1.
 DR PROSITE; PS00871; CLPAB_2; 1.
 DR Chaperone; ATP-binding; Repeat; Complete proteome.
 FT DOMAIN 161 409 I.
 FT NP_BIND 535 726 II.
 FT NP_BIND 206 213 ATP (POTENTIAL).
 FT NP_BIND 609 616 ATP (POTENTIAL).
 SQ SEQUENCE 878 AA; 99891 MW; 7D9E7419E42A2202 CRC64;
 Query Match 46.2%; Score 2104; DB 1; Length 878;
 Best Local Similarity 49.9%; Pred. No. 7.3e-85;
 Matches 434; Conservative 158; Mismatches 253; Indels 24; Gaps 7;
 Qy 1 MNEKFTKTNETIAFAHNAVAGHAQTPPLHLAGALISDPTGIPFQPAISSAGGNAQA 60
 Db 1 MNTDRYTVKASEALNDALSLAEAEHNGVVEEHLHALLSOKDGIISPLIEKIGAKPDF 60
 Qy 61 SARRVNOALKLPSQPPDDIPASSLLKIVRAQAQKSRGDTLHVLAVDOLIMGLLE- 119
 Db 61 YDE--LLQCLLRKPRVTGPAQTRCAPTLISKACARERLALKNQDEYVSCHELLLAIST 118
 Qy 120 DSQIRDLNVEGVATARKVSEKLRGKGGKVESASGDTNFQALKTYGRDLVEQA--GK 177
 Db 119 DSNTARLLHSQITSKTSIAALKDIRG--SKRVTSQDPSTFQCLEKCYCDLTTLAREK 176
 Qy 178 LDPVIGRDEIRRVRLISRTKNNPVLIGBPGVKTAVVGLAORIVKGVPSNLTDVR 237
 Db 177 IDEVIGRDEIRRVQVLSRRTKNNPVLIGBPGVKTAVVGLARRIVSGDVPESLKGKR 236
 Qy 238 LISLDGALVAGAKYRGFEERLKSVLKEVEDAEAGKVLFDIHLVLCAGKTEGMDAA 297
 Db 237 LLSLDLQALVAGAKRGFEERLKAVIDEIKGANGEVTFIDELHTIVGASGESMDAS 296
 Qy 298 NLFKPMARGQLRGICGATTLLEYKYVKEKDAAFRRRQQVYVAPSPVDTISILRLKKEK 357
 Db 297 NLKLPALARGELRSIGATTLNEYKYVKEKDAALERRRQQVYVCPVTVEDTIALRLQLEK 356
 Qy 358 YEGHGVRIQDRALINAQLSARVITGRHLDPKALDVIDEACANVRVOLDSPBEIDNLE 417
 Db 357 YEVHGVRIKDEALVAATVLSDRYITNTRFLPKAIDLVDDEAASRLKMETESQFVELDQVE 416
 Qy 418 RKEMQLEIEHLALEREKDKASKARLIEVRKELDDRLDKLOPLTMVKYKKEKIDEIRLK 477
 Db 417 RKLQNLNIEKASLLKESDPASKERLEKLEKELAGFLERRAAMQVQWQNEKGRIESRYYK 476

[illegible]

723 SNLGS-D-MIOLAEKSEEMKSAWVEIWAHFRDEFINRIDEAIVFHLAKTHYRIAQ 780
 779 LQMKDVAVLAEGRVALAVTDAALDYILAESVDVVGARPIRRWMEKKVVELSKWVRE 838
 781 IQLELRRLQTRLELLSVEEDAINQLVELGVDPLFGARPLKRAIQTNIENPLAQAALLAG 840

839 EIDENSTVVI 848

841 QVLPQSTITI 850

UNT 12

4 YEAST

H104 YEAST STANDARD; PRT; 908 AA.

P31539;

01-JUL-1993 (Rel. 26, Created)

01-FEB-1996 (Rel. 33, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Heat shock protein 104.

HSP104 OR YL026W OR L0948.

Saccharomyces cerevisiae (Baker's yeast).

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

NCBI_TaxID=4932;

[1]

SEQUENCE FROM N.A.

STRAIN=X21180 / ATCC 26109;

MEDLINE=91375541; PubMed=1896074;

Parsell D.A., Sanchez Y., Stitzel J.D., Lindquist S.;

"Hsp104 is a highly conserved protein with two essential nucleotide-

binding sites.";

Nature 353:270-273 (1991).

[2]

SEQUENCE FROM N.A.

STRAIN=S288c / AB972;

MEDLINE=97313267; PubMed=9169871;

Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Anserge W.,

Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,

Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,

Heuss-Neitzel D., Hilbert H., Hilger F., Klein K., Koetter P.,

Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,

Mueller-Auer S., Neutwich U., Obermaier B., Piravandi E., Pohl T.M.,

Portetalle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,

Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,

Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhaesselt P.,

Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,

Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;

"The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";

Nature 387:87-90 (1997).

[3]

SEQUENCE OF 749-908 FROM N.A.

STRAIN=S288c;

MEDLINE=97197984; PubMed=9046100;

Purnelle B., Goffeau A.;

"The sequence of 32kb on the left arm of yeast chromosome XII reveals

six known genes, a new member of the seripauperins family and a new

ABS transporter homologous to the human multidrug resistance

protein.";

Yeast 13:183-188 (1997).

!- FUNCTION: VITAL FOR TOLERANCE TO HEAT, ETHANOL AND OTHER STRESSES.

!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announcements>
 or send an email to license@isb.ch).

EMBL; M67479; AAA50477.1; -

DR EMBL; Z73131; CAA97475.1; -
 DR EMBL; Z73130; CAA97474.1; -
 DR EMBL; X97560; CAA66164.1; -
 DR PIR; S61476; S61476.
 DR SWISS-2DPAGE; F31539; YEAST.
 DR SGD; S0003949; HSP104.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0003767; F:co-chaperone activity; IDA.
 DR GO; GO:0003773; F:heat shock protein activity; IDA.
 DR GO; GO:0006457; P:protein folding; IDA.
 DR GO; GO:0006950; P:response to stress; IDA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003959; AAA_ATPase_cent.
 DR InterPro; IPR001270; Chaplin_clpA/B.
 DR InterPro; IPR004176; Clp_N.
 DR Pfam; PF00004; AAA; 1.
 DR Pfam; PF02861; Clp_N; 2.
 DR PRINTS; PR00300; CLIPROTEASEA.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00870; CLPAB 1; 1.
 DR PROSITE; PS00871; CLPAB 2; 1.
 KW Chaperone; Heat shock; ATP-binding; Repeat.
 FT DOMAIN 167 411
 II.
 FT NP BIND 541 731
 II.
 FT NP BIND 212 219 ATP (POTENTIAL).
 FT NP BIND 614 621 ATP (POTENTIAL).
 SQ SEQUENCE 908 AA; 102034 MW; 4AD0E7E3AF98E318 CRC64;
 Query Match 41.5%; Score 1890; DB 1; Length 908;
 Best Local Similarity 43.2%; Pred. No. 1.6e-75;
 Matches 405; Conservative 181; Mismatches 288; Indels 64; Gaps 16;
 QY 1 MNPE-KFTHTKNTIATATHELVNAGHAQFTPLHLAALISDP-TGIFPQAISAGGENA 58
 DB 1 MNDQTFERALTILTILTAQKLASDQHQPQLQPIHLAAFIETPDGSPYLV-----QNL 54
 QY 59 AQA-----ERVNQALKLPSQSPDDIPASSLIKIVIRRAQAQKSGDTHLAVD 111
 DB 55 IEKGRYDYDLFKVNVNLRVIRPQQPAPAEITPVALGKVLQDAKIQKQKQSFIAQD 114
 QY 112 QLIMGLSDSQIRDLNVEGVATATKVEKLRGKGVKVESASGDTN--FOALKTYGR 169
 DB 115 HILFALFNDSIQIFKEAQVDIEAKQALSLRG--NTRIDSRGADNTPLLEYLSKVAI 172
 QY 170 DLVQQA--GKLDVIGRDEIRRVVRIILSRRTKNNPVLITGEPGVGTAVVEGLAQIRVKG 227
 DB 173 DMTEQARQKGLDPVIGREBEIRSTIRVLARRIKSNPNFLIGBPGIKTAIEGVAQRIIDD 232
 QY 228 DYPNSLTVDRLISLDMGALVAGAKYRGFEERLKSVLKEVEDAEKGVILFIDEHLVLA 287
 DB 233 DVPTILQGAFLSLDLAALTAGAKYKGDPEERFKGLKEIEESKTLIVLFIIDEHLMGN 292
 QY 288 GKTEGSDAANLFPMLARGQRCIGATTLEIRYKVEKDAAFERRFOQVYVAEPSVPT 347
 DB 293 GKD----DAANILKALSEGQLKVIGATTNNEYSIVKDGAFERFQKIEVAEPSVRQT 348
 QY 348 ISILKGLKKEKYGHHGVRIQDRALNAAQLSARYITGRHLDPKALIDLVDKCANVVRVOLD 407
 DB 349 VAILRGLQPKYIEHHGVRIILDSALVTAQAQKLYPYRRLPDSALDLVDISCAGVAVARD 408
 QY 408 SOPERDNLERKEMQLELHAELEK--DKASKARLIEVRKELDLRLDKLQPLTMKYRK 465
 DB 409 SKEPEELDKERQLQIQVEIKALEDEDASTTKDLRLKARQKASLQEELEPLRQRYNE 468
 QY 466 EKERIDEIRLRKQKEELMFSLQEAERYDYLRARADLRAGIQAQVESAIAQEGTSSEE- 524
 DB 469 EKHGHEELTQAKKLDLENKALDAERYDYATAADLRVFAIPDIKKIEKLEDOVAEE 528
 QY 525 -----NVMJTEVNGPEHIAEVVVRWTGIPVTRLQNGEKERLIGLADRLHKRVVGQNAV 579
 DB 529 RRAGANSMTQNVVDTSITETAARLTGIPVKKLSESENEKLHMERDLSSSEVVGQMDA 588
 QY 580 AVSEALLSRAGLGRAQQTGSLFGTGTGKTELAKALAEQLFDDENLIVRDMSEYM 639

CLPB HELPY STANDARD; PRT; 856 AA.
 P71404;
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Clpb protein.
 CLPB OR HP0264.
 Helicobacter pylori (Campylobacter pylori).
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;

[1] SEQUENCE FROM N.A.
 STRAIN=NCTC 11637;
 Allan E.;
 Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

[2] SEQUENCE FROM N.A.
 STRAIN=26695 / ATCC 700392;
 MEDLINE=97394467; PubMed=9252185;
 Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
 Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 Hayes W.S., Borodovsky M., Kar P.D., Smith H.O., Fraser C.M.,
 Venter J.C.;
 "The complete genome sequence of the gastric pathogen Helicobacter
 pylori.";

Nature 388:538-547(1997).

!- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR

ATP-DEPENDENT PROTEASE (BY SIMILARITY).

!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; Y08238; CAA69406.1; -

EMBL; AE000545; AAD07330.1; -

PIR; H64552; H64552.

TIGR; HP0264; -

InterPro; IPR003593; AAA_Atpase.

InterPro; IPR003599; AAA_Atpase_cent.

InterPro; IPR001270; Chaprinin_clpb/B.

InterPro; IPR004176; Clp_N.

Pfam; PF00004; AAA; 2.

Pfam; PF02861; Clp_N; 2.

PRINTS; PR00300; CLPPTHEASEA.

SMART; SM00382; AAA; 2.

PROSITE; PS00870; CLPB_1; 1.

PROSITE; PS00871; CLPB_2; 1.

Chaperone; ATP-binding; Repeat; Complete proteome.

DOMAIN 159 406 II.

DOMAIN 531 722 II.

NP_BIND 204 211 ATP (POTENTIAL).

NP_BIND 605 612 ATP (POTENTIAL).

CONFLICT 14 14 T -> A (IN REF. 1).

CONFLICT 33 33 M -> L (IN REF. 1).

CONFLICT 59 59 Q -> E (IN REF. 1).

CONFLICT 62 62 R -> K (IN REF. 1).

CONFLICT 79 79 S -> N (IN REF. 1).

CONFLICT 103 103 R -> T (IN REF. 1).

CONFLICT 119 119 G -> S (IN REF. 1).

CONFLICT 131 131 A -> T (IN REF. 1).

CONFLICT 145 145 R -> A (IN REF. 1).

CONFLICT 149 149 D -> G (IN REF. 1).

FT	CONFLICT	222	222	M -> V (IN REF. 1).
FT	CONFLICT	435	435	A -> H (IN REF. 1).
FT	CONFLICT	465	465	V -> A (IN REF. 1).
FT	CONFLICT	643	644	MS -> IT (IN REF. 1).
FT	CONFLICT	649	649	A -> P (IN REF. 1).
FT	CONFLICT	734	734	E -> D (IN REF. 1).
FT	CONFLICT	765	765	D -> G (IN REF. 1).
FT	CONFLICT	836	836	V -> I (IN REF. 1).
FT	CONFLICT	840	840	D -> G (IN REF. 1).
SQ	SEQUENCE	856	AA; 96683	MM; E3902C7D989496AD CRC64;

Query Match 41.2%; Score 1876.5; DB 1; Length 856;
 Best Local Similarity 45.0%; Pred. No. 5.8e-75;
 Matches 390; Conservative 191; Mismatches 262; Indels 23; Gaps 11;

Qy	4	EKTHKNETIATHELANVAGHAQTPPLHAGALISDPTGIPPOAISAGGENAAQSAE	63
Db	5	EKMTDQLHETLSALALHKKNAEVTMHLFAMLNSQGILIQALQMPVD--IQALR	62
Qy	64	RVINOALKKLPSPSP--PPDIPASSSLIKVIRRAQAQKSRGDTHLAVDOLI--MGLLE	119
Db	63	LSVQSELNFKAKVQISQKNTQINQALIQSLENAQGLMAKRGDSFIATVYLLANGLFE	122
Qy	120	DSQIRDLNVEGVATARKSEVEKLRGKKGKVESASGDTNFQALKTYGRDVEQA--GK	177
Db	123	-SVLKPYLD----AKELQKTLBSLA--KGRTTQDKNDSDNLSLEKFGDILTKALENK	174
Qy	178	LDPVIGRDEIRVVRILSRPKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVNSLTVDR	237
Db	175	LDPVIGRDEIRVVRILSRPKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVNSLTVDR	234
Qy	238	LISLDGALVAGAKYRGEFEERLKSVLKVEVDAEGKLVILFIDEIHLVLGAGTSGMDAA	297
Db	235	VIALDLSLVAGAKYRGEFEERLKSVLKVEVDAEGKLVILFIDEIHLVLGAGTSGMDAA	293
Qy	298	NLPKPMARQRCICATTLLEKYRKYVEKDAAPERFQVYVAESPDPITISILGLKEK	357
Db	294	NILKPALARGELHTIGATTLLEKYRKYVEKDAAPERFQVYVAESPDPITISILGLKEK	353
Qy	358	YEGHGVQRDRLINAAQSLARYITGRHLPKATLDVDEACANVRVQLDSQPEIDNLE	417
Db	354	LETHNITINDSALIASAKLSRYITDFLPKADILIDEGAAQLKMQMSEPAKLISYK	413
Qy	418	RKMQLIEIHLALEREKDKASKARLFEVRKELDDLRDKLQPLTMKRYKEKIDETIRLK	477
Db	414	RSTQRLMEKQALEMEKESKAKMQEILKELSDLKKEKIQLEAQENEKEVKEISRLK	473
Qy	478	QKREELMPSLOEAREYDLARAADLYGAIQEVESAIAQLE--GTSSEENVMLTENVGPE	535
Db	474	MEMESLKGEAREFRNGDYQQAQGEIYKIPENKKEBELQKWEAMQOQALLQNALTE	533
Qy	536	-HTAEVSVRWGIPVTRLGONKEKELIGLADRLHKVVGQNAVNAVSAIILRSRAGLR	594
Db	534	NVIAEIVSQWTHIPVQKMLQSEKNRVINTSELOKRVVGQKAKAIKAIKAKRAGLSD	593
Qy	595	AQOQTGSLFLGPTGVGKTELAKALAEQLFDENLIVRIDMSEYMEQHSVSLIGAPGY	654
Db	594	SNKPIGSLFLGPTGVGKTESAKALAEQLFDSDKNLIRIDMSEYMEQHSVSLIGAPGY	653
Qy	655	VGHEGGQLTAVRRRPYCVILFDEVEKARHVAFTNLLQVLDGRLTGGQRTVDFRSV	714
Db	654	VGYEGGQLTAVRRRPYCVILFDEVEKARHVAFTNLLQVLDGRLTGGQRTVDFRSV	713
Qy	715	IIMTSNLGAHLLAGITGVKVTMEVARDVVRVVRKHPPELNLRLDEIVDFPLSHDLR	774
Db	714	LIITSNVAGALLEENLSEAKQKA---IKESLRQFFKPEFLNRLDEIISFALDSHVI	770
Qy	775	KVARLQMDVAVRLAERGVAVLTAADLYILAESYDYPYVGARPIRRWMEKKVVTLSKM	834
Db	771	NIVGILFENIQQKALERGINITLDBEAKELIAEAGDFRFGARPLKALYEWEDKLAEL	830
Qy	835	VREEIDENSTVYIDAGAGDIIVRYE	860

831 ILEKVKENDSVAFVVENNEIVPKIK 856

8 CANAL STANDARD; PRT; 812 AA.
 Q96UK5;
 15-SEP-2003 (Rel. 42, Created)
 15-SEP-2003 (Rel. 42, Last sequence update)
 15-SEP-2003 (Rel. 42, Last annotation update)
 Heat shock protein 78, mitochondrial precursor.
 HSP78.
 Candida albicans (Yeast).
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Mitosporic Saccharomycetales; Candida.
 NCBI TaxID=5476;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=SC5314;
 Bhattacherjee V., Hostetter M.;
 "The Candida albicans HSP78 gene encoding a member of the ClpB family
 of stress proteins.";
 Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: SEEMS TO BE INVOLVED IN THE TURNOVER OF UNASSEMBLED
 MITOCHONDRIAL PROTEINS (By similarity)
 -!- SUBUNIT: POSSIBLY PART OF AN ATP-DEPENDENT PROTEASE.
 -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch

 EMBL: AF399713; AAK97626.1;
 InterPro: IPR003593; AAA_Atpase.
 InterPro: IPR003959; AAA_Atpase_Centr.
 InterPro: IPR001270; Chaptrn1n_clpA/B.
 Pfam: PF00004; AAA; 1.
 PRINTS: PR00300; CLPPTROTEASEA.
 SMART: SMO0382; AAA; 2.
 PROSITE: PS00870; CLPAB 1; 1.
 PROSITE: PS00871; CLPAB 2; 1.
 Chapterone; ATP-binding; Repeat; Heat shock; Mitochondrion;
 Transit peptide.
 TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
 CHAIN ? 812 HEAT SHOCK PROTEIN 78.
 NP_BIND 133 140 ATP (POTENTIAL).
 NP_BIND 534 541 ATP (POTENTIAL).
 SEQUENCE 812 AA; 91680 MW; 4596001CA6DE7276 CRC64;

 Query Match 40.5%; Score 1844.5; DB 1; Length 812;
 Best Local Similarity 46.1%; Pred. No. 1:3e-73;
 Matches 382; Conservative 162; Mismatches 219; Indels 65; Gaps 10;
 117 LLEDSQIRDLN---EYGVATARK-----SEVEKLRGKE 148
 11 LKQTSATSVNLAKSIPITATRYRPNQYANELAKPNVFTIHNIPSPCFQVNRHSSF 70
 149 GKVV---ESASGDTNFOALKYGRDLVEQA--GKLDPIVIGRDEIRKVRNLSRRTKNNP 203
 71 PRKLOMOQTEGDD-NRPALEKFGSDDLQLAKEGKLDPIVIGRDEIRRTIQLSRRTKNNP 129
 204 VLIPEGVGKTAIVVEGLAQRIKGVDPNSLTDVRLISLDMGALVAGAKYGEFEELKSV 263
 130 VLIAGTGTATWEGLAQRIIRGEVDPNSMDQIITLDLAGIISGAKYRGDFESKLKSI 189
 264 LKEVEDAGKVLIFDIHVLGAGKTGSGMDANLFPKMLARGQLRCIGATTLEEVRYK 323
 190 LKEVEKNGKVLIFDFEHLMLGMLGAKSGSIDASNLKLPALARGKLSMCGATTIEYRYK 249

QY 324 VEKDAFERBFOQVYVAEPPVDTTISILRGLKEKYEKGHVRIODRALINAAQISARYIT 383
 DB 250 VEKDAALARRPSPVTVNEPTGDDTISILRGLKEKYEKGHVRIODRALINAAQISARYIT 309
 QY 384 GRHLPDKAIDLVDSEACANRVQLDSQPEEIDNLERKMOLEIEHALEREKDKASKARLI 443
 DB 310 DRFLPDKAIDLVDSEASTLRLQHESPDAITADQIMWTIEIELESURKEEDQLSIDRKH 369
 QY 444 EVRKELDDLRLKLOPLTMKYRKKEKIDEIRRLKQKRHELMFSLQEAERYDLARAADLR 503
 DB 370 KLEKELEVKKSELKELTDQWSESEKRAIDAVKNAKSELEKAKYLEQATREGDYARASRIQ 429
 QY 504 YGAIQEVESAIAOLEGT--SSEENVMLTENVGPEHIAEYVSRWTGIPVTVLQNEKERLI 561
 DB 430 YASIPLOQDKIQELSNNELHAKSSNLLHDSVTSEDIAGVISKMTGIPVNNLLKGEKDKLL 489
 QY 562 GLADRLHKRVVQGNQAVNAVSEAILRSRAGLGRAQOQPTGSLFLGPTGVGKTELAKALAE 621
 DB 490 DMNILLRQSVQDEAIDAVSDAVRLQAGLTSENPPIASFMLGPTGTGKTELTSIAQ 549
 QY 622 QLFDDENLLVRIDMEYMEQHSVSRILIGAPGVVGHGEGQLTEAVRRPYPYCVILFDEVE 681
 DB 550 FLFNDKNAVVRFDMSFQEKHTISRLIGSPPGVGYESEGELTEAVRRKPYSVVLFDEFE 609
 QY 682 KAHVAVENTILLOVLDGRLTDGQGRVDFRNSVIMTSLNGAEHLIAGLT----GKVTME 737
 DB 610 KAHFDLSKULLQVLDGSLTDSHGKIDFNIIWISNIGQELLADKNTYEDDGHINSE 669
 QY 738 VARDQVNRVHRHFRPELNLRLDEIVVFDPLSHDQLRKVARLQMKDVAVRLAERGVALAV 797
 DB 670 V-KSQVLENLRHHYAPEFLNRIDDIWVFNRLSKTALKEITDIRLEIGRLVDKRIILQL 728
 QY 798 TDAALDYILAESVDPVYGARPIRRWKEKVVTELKMWVREIDENSTVIIDAGAGDLVY 857
 DB 729 TDEAKTLLCDMGVDPTYGARPLNRVLRKLLDPLAMRLIKGQVQENETVKVEV----- 781
 QY 858 RVESGGLVDASTGKSDVLIHLIANGPKRSDDAAQAVKMKRIEIEDDN 905
 DB 782 -----KDHKIYV--PNHSEGTVIEKEBEDYFKEDKDDN 812

Search completed: February 13, 2004, 01:33:14
 Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 12, 2004, 22:31:20 ; Search time 73 Seconds

(without alignments)
3220.355 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

uence: 1 MNPEKFKHKTETATAHEL.....VKQRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 830525 seqs, 258052604 residues

al number of hits satisfying chosen parameters: 830525

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvivirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Alt	Score	Query Match	Length	DB ID	Description
1	4003	88.0	911	10	Q39889 glycine max
2	3896	85.6	909	10	Q9T113 nicotiana t
3	3883.5	85.4	912	10	Q98822 zea mays t
4	3840.5	84.4	912	10	Q8W2B5 oryza sativ
5	3825	84.1	918	10	Q9XEI1 oryza sativ
6	3822.5	84.0	913	10	Q9XEI1 oryza sativ
7	3807.5	83.7	912	10	Q8L6I4 oryza sativ
8	3790	83.3	918	10	Q9T112 oryza sativ
9	2502.5	55.0	582	10	Q9SY59 zea mays (m
10	2431	53.4	668	10	Q23323 arabidopsis
11	2299.5	50.5	871	16	Q8DJ40 synechococc
12	2298	50.5	867	5	O00828 leishmania
13	2294	50.4	931	16	Q8YJ91 bruceella me
14	2293	50.4	874	16	Q9AEM5 bruceella su
15	2289	50.3	867	5	Q25317 leishmania
16	2258.5	49.6	865	16	Q97KGO clostridium

17	2242.5	49.3	866	16	Q929G7
18	2238	49.2	866	16	Q8XKG8
19	2236.5	49.2	866	16	Q8Y570
20	2234.5	49.1	905	3	O94641
21	2225	48.9	868	16	Q98G96
22	2224	48.9	859	16	Q9A9T4
23	2218.5	48.8	835	16	Q8YM56
24	2217.5	48.7	883	2	O87444
25	2209	48.5	868	16	Q92MK7
26	2204	48.4	857	16	Q8XFM5
27	2199	48.3	887	16	Q8U8B5
28	2197	48.3	855	16	Q8CKC0
29	2185	48.0	889	16	O8G4X4
30	2184	48.0	852	16	Q8FM94
31	2182.5	48.0	898	16	P74459
32	2180.5	47.9	871	16	Q8PHQ4
33	2174.5	47.8	861	16	Q8P6A0
34	2168.5	47.7	858	16	Q92EA9
35	2166	47.6	968	10	O9LF37
36	2165.5	47.6	865	16	O8CJV9
37	2163.5	47.5	857	16	Q92JK8
38	2161	47.5	864	16	Q8D181
39	2160	47.5	857	16	Q8ZBV9
40	2158.5	47.4	861	16	Q9PGC1
41	2158.5	47.4	877	2	Q92AX2
42	2151	47.3	857	2	Q9KHC1
43	2150.5	47.3	848	16	Q9CB26
44	2144.5	47.1	864	16	Q8RHQ8
45	2141.5	47.1	869	16	Q99VB5

ALIGNMENTS

RESULT 1

ID	Q39889	PRELIMINARY;	PRT;	911 AA.
AC	Q39889;			
DT	01-NOV-1996 (TREMREL. 01, Created)			
DT	01-NOV-1996 (TREMREL. 01, Last sequence update)			
DT	01-JUN-2002 (TREMREL. 21, Last annotation update)			
DE	Heat shock protein.			
GN	SB100.			
OS	Glycine max (Soybean).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	euroside 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.			
OX	NCBI_TaxID=3847;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95170290; PubMed=7866031;			
RA	Lee Y.R., Nagao R.T., Key J.L.;			
RT	"A soybean 101-KD heat shock protein complements a yeast HSP104			
RL	Plant Cell 6:1889-1897(1994).			
DR	EMBL; L35272; AAA66338.1;			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003959; AAA_ATPase_cent.			
DR	InterPro; IPR001270; Chaprinin_cipa/B.			
DR	InterPro; IPR004176; Clp_N.			
DR	Pfam; PF02861; Clp_N; 2.			
DR	PRINTS; PR00300; CLP_PROTASEA.			
DR	SMART; SM00382; AAA; 2.			
DR	PROSITE; PS00870; CLPAB_1; 1.			
DR	PROSITE; PS00871; CLPAB_2; 1.			
DR	ATP-binding.			
SK	SEQUENCE 911 AA; 101328 MW; 6A2E1EE309067884 CRC64;			

Query Match 88.0%; Score 4003; DB 10; Length 911;
Best Local Similarity 87.0%; Pred. No. 1.3e-187;
Matches 794; Conservative 72; Mismatches 39; Indels 8; Gaps 6;

1 MNPEKFTHTNETTATATHELAHVAGHAQFTPLHLAGALISDPTGIFPQAISGA-GENAA 59
 1 MNPEKFTHTNETTATATHELAHVAGHAQFTPLHLAGALISDPTGIFPQAISGA-GENAA 60
 60 QSAERVINQALKKLPQSPPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 119
 61 RAVERVNLQALKKLPQSPPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 120
 120 DSQTRDLLNEGVATARYKSEVEKLRGKGGKVSASGDTNFQALTKYGRDLVQAGKLD 179
 121 DSQIGDLLKEAGVAVAKVSEVDKLRGKGGKVSASGDTNFQALTKYGRDLVQAGKLD 180
 180 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 181 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 240
 240 SLDGALVAGAKYRGEFEERLKSVLKEVEDEAGKVLFIIDEHLVLAGKTEGSDMAANL 299
 241 ALDMGALVAGAKYRGEFEERLKSVLKEVEDEAGKVLFIIDEHLVLAGKTEGSDMAANL 300
 300 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPSPDTISILRGLKERYE 359
 301 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPSPDTISILRGLKERYE 360
 360 GHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPQEIIDLK 419
 361 GHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPQEIIDLK 420
 420 RMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKK 479
 421 RMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKK 480
 480 REELMFSLOEABERYDLPADRYCAIQEVESAIAQLEGTSSEENVMLTENVGPEHIAE 539
 481 REELMFSLOEABERYDLPADRYCAIQEVESAIAQLEGTSSEENVMLTENVGPEHIAE 539
 540 VVSRWTGIPVTRLGONEKERLIGLADRLHVRVVGQNAVSEALIRSRAGLGRQOPT 599
 540 VVSRWTGIPVTRLGONEKERLIGLADRLHVRVVGQNAVSEALIRSRAGLGRQOPT 599
 600 GSFLFGPTGVGKTELAKALAEQFDENQLVRIDMSYMEQHSVRLIGAPGVVGHHE 659
 600 GSFLFGPTGVGKTELAKALAEQFDENQLVRIDMSYMEQHSVRLIGAPGVVGHHE 659
 660 GQGLTEAVRRPYCVILFDEVEKAVHVAFTLLQVLDGRLTDGQRTVDFNSVIMTS 719
 660 GQGLTEAVRRPYCVILFDEVEKAVHVAFTLLQVLDGRLTDGQRTVDFNSVIMTS 719
 720 NLGAELHLAGLTKVMEVARCVMEVRKFRPELLNRLDEIVVDFPLSHDQLEKVARL 779
 720 NLGAELHLAGLTKVMEVARCVMEVRKFRPELLNRLDEIVVDFPLSHDQLEKVARL 779
 780 QMKDVAVRLAERGVAVTDAALDYLAESYDPVYGARPIRMWKEKVVVTELSKRVVREE 839
 780 QMKDVAVRLAERGVAVTDAALDYLAESYDPVYGARPIRMWKEKVVVTELSKRVVREE 839
 840 IDENSTVYIDAG--AGDLVRYVE--SGGLVDASTGKSDVILHIANG--PKSDAAQAVK 894
 840 IDENSTVYIDAG--AGDLVRYVE--SGGLVDASTGKSDVILHIANG--PKSDAAQAVK 894
 895 MRIEIEDDDNEE 907
 899 MKIEIEDDDNEE 911

SULT 2
 Q92T13
 Q92T13
 01-MAY-1999 (TReMBLrel. 10, Created)
 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 101 kDa heat shock protein.

GN HSP101.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
 OC NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9903076; PubMed=9784498;
 RA Wells D.R., Tanguay R.L., Le H., Gallie D.R.;
 RT "HSP101 functions as a specific translational regulatory protein whose
 activity is regulated by nutrient status.";
 RL Genes Dev. 12:3236-3251(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gallie D.R., Tanguay R.L., Wells D.R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF083343; AAC83688.2;
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001270; Chaprinin_c1pA/B.
 DR InterPro; IPR004176; C1p_N.
 DR Pfam; PF00004; AAA; 2.
 DR Pfam; PF02861; C1p_N; 2.
 DR PRINTS; PR00300; CLPPTTEASEA.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00870; CLPAB_1; 1.
 DR PROSITE; PS00871; CLPAB_2; 1.
 KW ATP-binding; Heat shock.
 SQ SEQUENCE 909 AA; 101110 MW; 86641AD7FA47678D CRC64;

Query Match 85.6%; Score 3896; DB 10; Length 909;
 Best Local Similarity 85.0%; Pred. No. 2.3e-182;
 Matches 774; Conservative 75; Mismatches 56; Indels 6; Gaps 5;

QY 1 MNPEKFTHTNETTATATHELAHVAGHAQFTPLHLAGALISDPTGIFPQAISGA-ENAA 59
 DB 1 MNPEKFTHTNETTATATHELAHVAGHAQFTPLHLAGALISDPTGIFPQAISGA-ENAA 60
 QY 60 QSAERVINQALKKLPQSPPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 119
 DB 61 NSERVNLQALKKLPQSPPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 120
 QY 120 DSQTRDLLNEGVATARYKSEVEKLRGKGGKVSASGDTNFQALTKYGRDLVQAGKLD 179
 DB 121 DSQIGDLLKEAGVAVAKVSEVDKLRGKGGKVSASGDTNFQALTKYGRDLVQAGKLD 180
 QY 180 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 DB 181 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 240
 QY 240 SLDGALVAGAKYRGEFEERLKSVLKEVEDEAGKVLFIIDEHLVLAGKTEGSDMAANL 299
 DB 241 ALDMGALVAGAKYRGEFEERLKSVLKEVEDEAGKVLFIIDEHLVLAGKTEGSDMAANL 300
 QY 300 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPSPDTISILRGLKERYE 359
 DB 301 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPSPDTISILRGLKERYE 360
 QY 360 GHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPQEIIDLK 419
 DB 361 GHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPQEIIDLK 420
 QY 420 RMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKK 479
 DB 421 RMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKK 480
 QY 480 REELMFSLOEABERYDLPADRYCAIQEVESAIAQLEGTSSEENVMLTENVGPEHIAE 539
 DB 481 REELMFSLOEABERYDLPADRYCAIQEVESAIAQLEGTSSEENVMLTENVGPEHIAE 539
 QY 540 VVSRWTGIPVTRLGONEKERLIGLADRLHVRVVGQNAVSEALIRSRAGLGRQOPT 599
 DB 540 VVSRWTGIPVTRLGONEKERLIGLADRLHVRVVGQNAVSEALIRSRAGLGRQOPT 599

540 VVSRWTGIPVSRIGQNEKEKLGIDRLHQRVVGQDHAVRAVAEAVLSRAGLRPOOPT 599
600 GSFLPGTGVGKTELAKALAEQLEDDENLLVRIDMSEYMSHSRSLIGAPPGVVGHEE 659
600 GSFLPGTGVGKTELAKALAEQLEDDENLLVRIDMSEYMSHSRSLIGAPPGVVGHEE 659
660 GQGLTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQGRVDFRNSVIIMTS 719
660 GQGLTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQGRVDFRNSVIIMTS 719
720 NLGAELHLAGLTKVMEYARDVCMVRKHPPELLNRLDEIVVDFPLSHDQRLKVARL 779
720 NLGAELHLAGLTKVMEYARDVCMVRKHPPELLNRLDEIVVDFPLSHDQRLKVARL 779
780 QMKDVAVRLAERGVALAVTDAALDVLAEVDYVYGARPIRRMMEKKVVTLSKVVVREE 839
780 QMKDVAVRLAERGVALAVTDAALDVLAEVDYVYGARPIRRMMEKKVVTLSKVVVREE 839
840 IDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKOR 896
840 IDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKOR 896
897 IEIEDDDNEE 907
899 IEIEDDDMED 909

UT 3
8822
Q98822 PRELIMINARY; PRT; 912 AA.
Q98822; 01-MAY-2000 (TremBrel. 13, Created)
01-MAY-2000 (TremBrel. 13, Last sequence update)
01-JUN-2002 (TremBrel. 21, Last annotation update)
Heat shock protein HSP101.
HSP101.
Zea mays (Maize).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
PACCAD clade; Panicoidae; Andropogoneae; Zea.
NCBI_TaxID=4577;
(1)
SEQUENCE FROM N.A.
Young T.E., Gallie D.R.;
"HSP101 Diversity in Monocot Species";
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
(2)
SEQUENCE FROM N.A.
MEDLINE=99234099; PubMed=10216257;
Nieto-Sotelo J., Kannan K.B., Martinez L.M., Segal C.;
"Characterization of a maize heat shock protein 101 gene, HSP101,
encoding a ClpB/Hsp100 protein homologue";
Gene 230:187-195 (1999).
EMBL; AF133840; AAD33606.1; -;
EMBL; AF077337; AAD25223.1; -;
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_cent.
InterPro; IPR001270; Chaprinin_c1pA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 2.
PRINTS; PR00300; CLPPTROTEASE.
SMART; SM00382; AAA; 2.
PROSITE; PS00870; CLPAB_1; 1.
PROSITE; PS00871; CLPAB_2; 1.
ATP-binding; Heat shock.
SEQUENCE 912 AA; 13FIEAA4BEA610E2 CRC64;
very Match 85.4%; Score 3893.5; DB 10; Length 912;
Best Local Similarity 84.6%; Pred. No. 9.4e-182;
Matches 772; Conservative 76; Mismatches 58; Indels 7; Gaps 6;
1 MNPEKTHKNTIATAHAVNAGHQAQTPLHLAALISDPTGTFPPQAISGA-GENNA 59

1 MNPEKTHKNTIATAHAVNAGHQAQTPLHLAALISDPTGTFPPQAISGA-GENNA 60
60 -QSAERVINQALKKLPSPSPDDIPASSSLIKVIRRAQAQKSRGTHLAVDQIMGLL 118
61 GDSFVRVNNLSKKLPSPSPDDIPASSSLIKVIRRAQAQKSRGTHLAVDQIMGLL 120
119 EDSQIRDLNVEGVATARKVEKELGEGKGVESAGDTNFQALKTGYGRDLVEQAGKL 178
121 EDSQISCLKEAGVSAARVARELEKRGEGRRVESAGDTNFQALKTGYGRDLVEQAGKL 180
179 DPVIGRDEIRRVVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVNSLTDVRL 238
181 DPVIGRDEIRRVVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVNSLTDVRL 240
239 ISLDMGALVAGAKYGFEEERLKSIVKEVEDAGKVLFIIDEHLVLGAGKTEGSDAAN 298
241 IALDMGALVAGAKYGFEEERLKSIVKEVEDAGKVLFIIDEHLVLGAGKTEGSDAAN 300
299 LFKPMLARGQLRCIGATTLEEVRYKVEKDAAPERRFQVYVAEPSPDTISILRGKEKY 358
301 LFKPMLARGQLRCIGATTLEEVRYKVEKDAAPERRFQVYVAEPSPDTISILRGKEKY 360
359 EGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEEIDNLER 418
361 EGHGVRIQDRALVAAQLSARYIMGRHLPDKAIDLVDACANVRVQLDSQPEEIDNLER 420
419 KQMLELHALEREKDKASKARLIEVRKELDDLDKLOPLTMKYRKEKERIDEIRRLKQ 478
421 KRIQLEVLHLEKEKDKASKARLIEVRKELDDLDKLOPLTMKYRKEKERIDEIRRLKQ 480
479 KREELMFSIQEAERYRLARAADLRVGAIOEVESAIAQLEGTSSSEENYMLTENVPHEITA 538
481 RREELQFTLQEAERRMDLARVADLYKGLAQLEDAISKLE-SETGENMLTETVGEPIA 539
539 EVVSRWTGIPVTRLGQNEKERLIGLADRLHQRVVGQDHAVRAVAEAVLSRAGLRPOOPT 598
540 EVVSRWTGIPVTRLGQNEKERLIGLADRLHQRVVGQDHAVRAVAEAVLSRAGLRPOOPT 599
599 TGSFLPGTGVGKTELAKALAEQLEDDENLLVRIDMSEYMSHSRSLIGAPPGVVGHEE 658
600 TGSFLPGTGVGKTELAKALAEQLEDDENLLVRIDMSEYMSHSRSLIGAPPGVVGHEE 659
659 EGGQTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQGRVDFRNSVIIMT 718
660 EGGQTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQGRVDFRNSVIIMT 719
719 SNLGAELHLAGLTKVMEYARDVCMVRKHPPELLNRLDEIVVDFPLSHDQRLKVAR 778
720 SNLGAELHLAGLTKVMEYARDVCMVRKHPPELLNRLDEIVVDFPLSHDQRLKVAR 779
779 LQMKDVAVRLAERGVALAVTDAALDVLAEVDYVYGARPIRRMMEKKVVTLSKVVVREE 838
780 LQMKDVAVRLAERGVALAVTDAALDVLAEVDYVYGARPIRRMMEKKVVTLSKVVVREE 839
839 EIDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKOR 895
840 EIDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKOR 896
896 RI-IEIEDDDNEE 907
900 RIMEEDDDMED 912

RESULT 4
Q98285 PRELIMINARY; PRT; 912 AA.
ID Q98285
AC Q98285;
DT 01-MAR-2002 (TremBrel. 20, Created)
DT 01-MAR-2002 (TremBrel. 20, Last sequence update)
DT 01-OCT-2002 (TremBrel. 22, Last annotation update)
DE Heat shock protein.
GN HSP101.

Oryza sativa (Rice).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Rhizophytidae; Oryzoideae; Oryzae; Oryza.

NCBI_TaxID=4530;

SEQUENCE FROM N.A.
STRAIN=cv. Nipponbare;
Agarwal M., Sahi C., Young T., Agarwal S., Mitra S., Ganesan K.,
Agarwal-Katihar S., Gallie D.R., Grover A.,
"Isolation and Analysis of HSP101 from Rice."
Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF332981; AAL57165.1; -

Gramene; Q8W2B5;
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_cent.
InterPro; IPR001270; Chaprinin_clpA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
PRINTS; PR00300; CLP_PTEASEA.
SMART; SM00382; AAA; 2.
PROSITE; PS00870; CLPAB_1; 1.
PROSITE; PS00871; CLPAB_2; 1.
ATP-binding.

SEQUENCE 912 AA; 100896 MW; 0A3B17065B7CA343 CRC64;

Query Match 84.4%; Score 3840.5; DB 10; Length 912;

Best Local Similarity 83.8%; Pred. No. 1.2e-179;

Matches 765; Conservative 75; Mismatches 66; Indels 7; Gaps 5;

1 MNPKEFTHTKNTTATATHELVANAGHAQFTPLHLAAGALISDPTGIFPQAISSAGGENAA- 59
1 MNPONFTHTKNEALVAHEIASEAGHAQLTPLHLVAALADKGGILFQAISSQASGGDAGA 60
60 -QSAERVINOALKLPQSPPDDIPASSSLIKVIRRAQAQKSRGTHLAVDQLMGLL 118
61 PDSFVVGALKKLPQSPPDDIPASTALIKVIRRAQAQKSRGTHLAVDQLMGLL 120
119 EDSQIRDLNVEGVATVRSEKLGKGEKVESAGDTNFQALTKYGRDVEQAGKL 178
121 EDSLISDLCEAGVSAARVRAELKRGEGKVESAGDTNFQALTKYGRDVEQAGKL 180
179 DPVIGRDETRVRVILSRRTKNNPVLIGEGVGTAVGGLAQIRIVKGVDPNSLTDVRL 238
181 DPVIGRDETRVRVILSRRTKNNPVLIGEGVGTAVGGLAQIRIVKGVDPNSLTDVRL 240
239 ISLDMGALVAGAKYRGEFEERLKSVEDEDAEGKVLFTDEIHLVLAGKTRGSDMAAN 298
241 ISLDMGALVAGAKYRGEFEERLKSVEDEDAEGKVLFTDEIHLVLAGKTRGSDMAAN 300
299 LFKPLMARGQLRCIGATTLEBYRYKYVEKDAAFERRFQVYVAEPSVPTTISILRGLKEKY 358
301 LFKPLMARGQLRCIGATTLEBYRYKYVEKDAAFERRFQVYVAEPSVPTTISILRGLKEKY 360
359 EGHGVRIQDRALNAAQLSARYITGRHLPKADLDLVEACANRVQLDSQPEIDMLER 418
361 EGHGVRIQDRALNAAQLSARYITGRHLPKADLDLVEACANRVQLDSQPEIDMLER 420
419 KRMQLEIHLALEREKDKASKARLIEVREKDLDDRLDKLPMTMYRKEKEIRDLRLKQ 478
421 KRIQLEVEHHALEKOKASARLVVEKDELDLDDRLDKLPMTMYRKEKEIRDLRLKQ 480
479 KRELMFSLQBAERYDARADLYRGAIEVESAIAQLEGTSEENVMLTENVGPEHIA 538
481 KREELQFTQBAERMOLARVADLYKALQEIADVAIAKLE-SETGENMLTETVGPQIA 539
539 EVSWRTGIPVTRIGONKEKRLIGLADLRHVRVGVQGNVAVSEAILRSAGLGRAQOP 598
540 EVSWRTGIPVTRIGONKEKRLIGLADLRHVRVGVQGNVAVSEAILRSAGLGRAQOP 599
599 TGSFLEFLGTGVTGKTELAKALAEQFDDENLLVRIDMSYMEQHSVRLIGAPPGYVGE 658

Db 600 TGSFLEFLGTGVTGKTELAKALAEQFDDENLLVRIDMSYMEQHSVRLIGAPPGYVGE 659
QY 659 EGGQTEAVRREPCYVILFDEVEKKAHVAVENTLLQVLDGRLTQGGRTVDFRNSVIIMT 718
Db 660 EGGQTEAVRREPCYVILFDEVEKKAHVAVENTLLQVLDGRLTQGGRTVDFRNSVIIMT 719
QY 719 SNLGAELHLAGTGTGKTEVAVRDCVMREVRKHFPELLNRLDEIVVFDPLSHDQRLKVAR 778
Db 720 SNLGAELHLAGTGTGKTEVAVRDCVMREVRKHFPELLNRLDEIVVFDPLSHDQRLKVAR 779
QY 779 LOMKDVAVRLAERGVALATDAALDYILAESYDVPYVGARPIRRWEKKVVTLSKVVRE 838
Db 780 LOMKDVAVRLAERGVALATDAALDYILAESYDVPYVGARPIRRWEKKVVTLSKVVRE 839
QY 839 EIDENSTVYIDAG--AGDLVVRVES--GGLVDASTGKSDVLIHIANGPKR--SDAAQAVKK 894
Db 840 EIDENSTVYIDAG--AGDLVVRVES--GGLVDASTGKSDVLIHIANGPKR--SDAAQAVKK 899
QY 895 MRIEIEDDDNEE 907
Db 900 MRIMEDEDDGMDDE 912

RESULT 5

Q9XE11 PRELIMINARY; PRT; 918 AA.
AC Q9XE11;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Heat shock protein 101.
GN HSP101A.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
[1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Mustang;
RX MEDLINE=21240210; PubMed=11342108;
RA Campbell J.I., Klueva N.Y., Zheng H.G., Nieto-Sotelo J., Ho T.H.D.,
Nguyen H.T.;
RT "Cloning of new members of heat shock protein HSP101 gene family in
RT wheat (Triticum aestivum (L.) Moench) inducible by heat, dehydration,
RT and ABA";
RL Biochim. Biophys. Acta 1517:270-277 (2001).
RL EMBL; AF097363; RAD22629.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaprinin_clpA/B.
DR InterPro; IPR004176; Clp_N.
DR Pfam; PF00004; AAA; 2.
DR PRINTS; PR00300; CLP_PTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding.
SQ SEQUENCE 918 AA; 101122 MW; 354D3BAC06C8556C CRC64;

Query Match 84.1%; Score 3825; DB 10; Length 918;

Best Local Similarity 82.6%; Pred. No. 6.9e-179;

Matches 760; Conservative 82; Mismatches 66; Indels 12; Gaps 7;

QY 1 MNPKEFTHTKNTTATATHELVANAGHAQFTPLHLAAGALISDPTGIFPQAISSAGGENAA- 59
Db 1 MNPONFTHTKNEALVAHEIASEAGHAQLTPLHLVAALADKGGILFQAISSQASGGDAGA 60
QY 60 -QSAERVINOALKLPQSPPDDIPASSSLIKVIRRAQAQKSRGTHLAVDQLMGLL 118
Db 61 PDSFVVGALKKLPQSPPDDIPASTALIKVIRRAQAQKSRGTHLAVDQLMGLL 120

119 EDSQIDLLNEGVATARKSEVEKLRGKE-GKVESASGDTNFQALTKYGRDLVEQAGK 177
 121 EDAQADCLKEAGVSASRVRAELEKLRGGDNKRKVESAGDTNFQALTKYGRDLVEVAGK 180
 178 LDPVIGRDEIRRVIRLGRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVPSNLTQVR 237
 181 LDPVIGRDEIRRVIRLGRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVPSNLTQVR 240
 238 LISLDMGALVAGAKYRGEFEERLKSVEDEAGKVILFIDEIHLVLGAGTEGSMDDAA 297
 241 LVALDMGALVAGAKYRGEFEERLKAIVKEVEAGKVILFIDEIHLVLGAGTEGSMDDAA 300
 298 NLFKPMALGQLRCIGATTLEBYRKYVEKDAFERFQOYVVAEPSPVDTISILGLKEK 357
 301 NLFKPMALGQLRCIGATTLEBYRKYVEKDAFERFQOYVVAEPSPVDTISILGLKEK 360
 358 YEGHGVRIQDRALINAAQLSARYITGRHLPKADLVDDEACANVRVOLDSPSEIDNLE 417
 361 YEGHGVRIQDRALINAAQLSARYITGRHLPKADLVDDEACANVRVOLDSPSEIDNLE 420
 418 RKMQLIEIHALEREKDKASKARLIEVKEKELDDRLDKLQPLTKMYRKEKERIDEIRLK 477
 421 RKRIOLEVEIHALEREKDKASKARLIEVKEKELDDRLDKLQPLTKMYRKEKERIDEIRLK 480
 478 QKREELMFSLOAERYDILARAADLYGAIQEVESALQLEGTSEENVMLTENVGPEHI 537
 481 QRREELQFTLOAERYDILARAADLYGAIQEVESALQLEGTSEENVMLTENVGPEHI 539
 538 AEVVSRTWGIPTVRLGQNEKERLIGLADRLHRRVVGONQAVNAVEAILRSRAGLGRPQ 597
 540 AEVVSRTWGIPTVRLGQNEKERLIGLADRLHRRVVGONQAVNAVEAILRSRAGLGRPQ 599
 598 PTGSFLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 657
 600 PTGSFLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 659
 658 EGGQITEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQGTVDPRNSVIIMT 717
 660 EGGQITEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQGTVDPRNSVIIMT 719
 718 TSNLGAHLLAGITGKVTMEVARDVYRVRHFRPELLNRLDEIVFDPGLSHDQLRKA 777
 720 TSNLGAHLLAGITGKVTMEVARDVYRVRHFRPELLNRLDEIVFDPGLSHDQLRKA 779
 778 RLQKMDVAVLAERGVALAVTDAALDYILAESVDPVYGARPIRRMMKKVTELSSQVVR 837
 780 RLQKMDVAVLAERGVALAVTDAALDYILAESVDPVYGARPIRRMMKKVTELSSQVVR 839
 838 EETIDENSTVYIDAGD-LVYRVE-SGGLVDASTGKSDVLIHIANGP---KRSDAQA 891
 840 EETIDENSTVYIDAGD-LVYRVE-SGGLVDASTGKSDVLIHIANGP---KRSDAQA 899
 892 VKQWRIEEDDDNEMIED 911
 900 VKQWRIEEDDDNEMIED 917

Q9SPH4 PRELIMINARY; PRT; 913 AA.
 Q9SPH4
 01-MAY-2000 (TrEMBLrel. 13, Created)
 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 Heat shock protein 101.
 HSP101.
 Triticum aestivum (wheat).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 Triticeae; Triticum.
 NCBI TaxID=4565;
 [1]
 SEQUENCE FROM N.A.

STRAIN=cv. Mustang;
 MEDLINE=21240210; PubMed=11342108;
 RA Campbell J.L., Klueva N.Y., Zheng H.G., Nieto-Sotelo J., Ho T.H.D.,
 Nguyen H.T.;
 RT "Cloning of new members of heat shock protein HSP101 gene family in
 wheat (*Triticum aestivum* (L.) Moench) inducible by heat, dehydration,
 and ABA";
 RL Biochim. Biophys. Acta 1517:270-277 (2001).
 DR EMBL; AF174433; AAF01280.1; -;
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003599; AAA_ATPase_cent.
 DR InterPro; IPR001270; Chaprinin_c1pA/B.
 DR InterPro; IPR004176; Clp_N.
 DR Pfam; PF00004; AAA_2.
 DR Pfam; PF02861; Clp_N; 2.
 DR PRINTS; PR00300; CUPPROTEASEA.
 DR SMART; SMO0382; AAA; 2.
 DR PROSITE; PS00870; CLPAB 1; 1.
 DR PROSITE; PS00871; CLPAB 2; 1.
 KW ATP-binding.
 SQ SEQUENCE 913 AA; 100897 MW; 867E2EF137C8E7DB CRC64;

Query Match 84.0%; Score 3822.5; DB 10; Length 913;
 Best Local Similarity 83.3%; Pred. No. 9e-179;
 Matches 761; Conservative 75; Mismatches 69; Indels 9; Gaps 6;

QY 1 MNPKEPTHKNTTATTAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISSAGGEN--A 58
 DB 1 MNPDKFTHKNTTATTAHELAVNAGHAQFTPLHLAALAAADRSGILFQAIHAASGGNDAA 60
 QY 59 AQSARVINOALKKLPSQSPPPDDIPASSSLIKVIRRAQAQSRGDTHLAVDOLIMGLL 118
 DB 61 AESFERSVAAALKKLPSPSPPPDTPVASTSLVAVRAQAQSRGDTHLAVDOLIMGLL 120
 QY 119 EDSQIDLLNEGVATARKSEVEKLRGKEKGVESAGDTNFQALTKYGRDLVEQAGK 178
 DB 121 EDPOISDALKEAGISAARVKAEEKLRGGDNKRKVESAGDTNFQALTKYGRDLVEVAGK 180
 QY 179 DPVIGRDEIRRVIRLGRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVPSNLTQVR 238
 DB 181 DPVIGRDEIRRVIRLGRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVPSNLTQVR 240
 QY 239 ISLDMGALVAGAKYRGEFEERLKSVEDEAGKVILFIDEIHLVLGAGTEGSMDDAA 298
 DB 241 VALDMGALVAGAKYRGEFEERLKAIVKEVEAGKVILFIDEIHLVLGAGTEGSMDDAA 300
 QY 299 LFKPMLARGQLRCIGATTLEBYRKYVEKDAFERFQOYVVAEPSPVDTISILGLKEKY 358
 DB 301 LFKPMLARGQLRCIGATTLEBYRKYVEKDAFERFQOYVVAEPSPVDTISILGLKEKY 360
 QY 359 EGHGVRIQDRALINAAQLSARYITGRHLPKADLVDDEACANVRVOLDSPSEIDNLER 418
 DB 361 EGHGVRIQDRALINAAQLSARYITGRHLPKADLVDDEACANVRVOLDSPSEIDNLER 420
 QY 419 KRMQLIEIHALEREKDKASKARLIEVKEKELDDRLDKLQPLTKMYRKEKERIDEIRLK 478
 DB 421 KRIQLEVEIHALEREKDKASKARLIEVKEKELDDRLDKLQPLTKMYRKEKERIDEIRLK 480
 QY 479 KREELMFSLOAERYDILARAADLYGAIQEVESALQLEGTSEENVMLTENVGPHIA 538
 DB 481 KREELQFTLOAERYDILARAADLYGAIQEVESALQLEGTSEENVMLTENVGPHIA 539
 QY 539 EVSRTWGIPTVRLGQNEKERLIGLADRLHRRVVGONQAVNAVEAILRSRAGLGRPQ 598
 DB 540 EVSRTWGIPTVRLGQNEKERLIGLADRLHRRVVGONQAVNAVEAILRSRAGLGRPQ 599
 QY 599 TGSFPLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 658
 DB 600 TGSFPLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 659
 QY 659 EGGQITEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQGTVDPRNSVIIMT 718
 DB 660 EGGQITEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQGTVDPRNSVIIMT 719

719 SNLGAHLIAGLTGKVTMEVARDVMBREVRKHFPELNLRLDEIVFDPDLSHDQLRKVAR 778
 720 SNLGAHLIAGLVGN-SMKVARDLVMOEVRHFRPELNLRLDEIVFDPDLSHDQLRKVAR 778
 779 LQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKVVTELSKMWVRE 838
 779 LQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKVVTELSKMWVRE 838
 839 EIDENSTVIYIDAGAG--DLVYRVES--GGLVDASTGKSDVLIHIANPKRSDAAQAVKKM 895
 839 EIDENSTVIYIDAGAG--DLVYRVES--GGLVDASTGKSDVLIHIANPKRSDAAQAVKKM 895
 896 RI--EETEDDDNEE 907
 899 KIMQDSGEVDDME 912

Q8L614 PRELIMINARY; PRT; 912 AA.
 Q8L614;
 01-OCT-2002 (TrEMBLrel. 22, Created)
 01-MAR-2003 (TrEMBLrel. 22, Last sequence update)
 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 Heat shock protein 101.
 HSP101.
 Oryza sativa (Rice).
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 NCBI_TaxID=4530;
 [1]

SEQUENCE FROM N.A.
 STRAIN=cv. Nipponbare;
 Agarwal M., Sahi C., Young T., Agarwal S., Mitra S., Ganeshan K.,
 Katiyar-Agarwal S., Gallie D.R., Grover A.;
 "Isolation and analysis of HSP101 from rice";
 Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 EMBL: AJ316025; CAC87117.1; -
 Gramene; Q8L614; -
 InterPro; IPR003593; AAA ATPase.
 InterPro; IPR003959; AAA ATPase centr.
 InterPro; IPR001270; Chaprinin_c1pA/B.
 InterPro; IPR004176; Clp_N.
 Pfam; PF00004; AAA; 2.
 Pfam; PF02861; Clp_N; 2.
 PRINTS; PR00300; ClpPROTEASEA.
 SMART; SM00382; AAA; 2.
 PROSITE; PS00870; CLPAB_1; 1.
 PROSITE; PS00871; CLPAB_2; 1.
 ATP-binding.
 SEQUENCE 912 AA; 100836 MW; 07EA168AED0605C1 CRC64;

Query Match 83.7%; Score 3807.5; DB 10; Length 912;
 Best Local Similarity 83.2%; Pred. No. 4.9e-178;
 Matches 760; Conservative 77; Mismatches 69; Indels 7; Gaps 5;
 1 MNPEKTHKNTIATAHELAVNAGHAQFTPLHLAGALISDPTGIPQAISSAGGENAA- 59
 1 MPDNPFTKNEALVAEHAIEASEAGHAQFTPLHVAALAAADKGGILRQAISSAGGDAGA 60
 60 -QSAERVINQALKPLPSQPPDDIPASSLLKIVRRACAKSGDTHLAVDQIMGLL 118
 61 PDSFERYVSGALKPLPSQPPDDIPASTALIKIVRRACAKSGDTHLAVDQIMGLL 120
 119 EDSQTRDLLNEVGATARYKSEVEKLRGEGKKEKVSASGDTNFQALTYGRDLVEQAGKL 178
 121 EDSLISDCLKEAGVSAARVRAELEKLRGEGKKEKVSASGDTNFQALTYGRDLVEQAGKL 180
 179 DPVIGRDEIRRVILSRRTKNNPVILGEPGVGKTA VVEGLAQRIKVGDPVNSLTDVRL 238
 181 DPVIGRDEIRRVILSRRTKNNPVILGEPGVGKTA VVEGLAQRIKVGDPVNSLTDVRL 240

QY 239 ISIDMGALVAGAKYGEFEERLKSVLKEVEDAGKVVILFIDBIHLVLAGKTEGSDMAAN 298
 DB 241 IALDGMALVAGAKYGEFEERLKSVLKEVEDAGKVVILFIDBIHLVLAGKTEGSDMAAN 300
 QY 299 LFKPMLARGQLRCIGATTILEEYRKVVEKDAAFERFQOVYVAEPSVPTTISTLRGLKKEY 358
 DB 301 LFKPMLARGQLRCIGATTILEEYRKVVEKDAAFERFQOVYVAEPSVPTTISTLRGLKKEY 360
 QY 359 EGHGVRITODRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEEIDNLER 418
 DB 361 EGHGVRITODRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEEIDNLER 420
 QY 419 KRMQLEIELHALEREKDKASKARLEVEKELDLDLKLKLOPLTMKYRKEKERIDEIRLKQ 478
 DB 421 KRIQLEVEHHALEKDKASKARLEVEKELDLDLKLKLOPLTMKYRKEKERIDEIRLKQ 480
 QY 479 KREELMFSLOEARRDYDLARAADLYGAIQEVESAIAOLEGTSSEENVMTENYVGEHIA 538
 DB 481 RREELQFTLQEAERRMDLARVADLYGALQIDVIAKLE-SETGENMLTETVGPQIA 539
 QY 539 EVVSRWTGIPVTRIGONEKERLIGLADRLHVRVVGQNAVNAVSEAILRSRAGLGRQOP 598
 DB 540 EVVSRWTGIPVTRIGONEKERLIGLADRLHVRVVGQNAVNAVSEAILRSRAGLGRQOP 599
 QY 599 TGSFLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVSRLLGAPGYVGE 658
 DB 600 TGSFLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVSRLLGAPGYVGE 659
 QY 659 EGGQITAVRRPYCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDPRNSVLIIT 718
 DB 660 EGGQITAVRRPYCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDPRNSVLIIT 719
 QY 719 SNLGAHLIAGLTGKVTMEVARDVMBREVRKHFPELNLRLDEIVFDPDLSHDQLRKVAR 778
 DB 720 SNLGAHLIAGLVGN-SMKVARDLVMOEVRHFRPELNLRLDEIVFDPDLSHDQLRKVAR 779
 QY 779 LQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKVVTELSKMWVRE 838
 DB 780 LQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKVVTELSKMWVRE 839
 QY 839 EIDENSTVIYIDAG--AGDLVYRVES--GGLVDASTGKSDVLIHIANPKR-SDAAQAVKK 894
 DB 840 EIDENCTVIYIDAAAPHKDELAYRVNDRGLVNAETGQKSDILIQVPNGAATGSDAAQAVKK 899
 QY 895 MRIEEIEDDNEE 907
 DB 900 MRIMEDEDGMDDE 912

RESULT 8
 Q9ZT12
 ID Q9ZT12 PRELIMINARY; PRT; 918 AA.
 AC Q9ZT12;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE 101 kDa heat shock protein.
 GN HSP101.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP MEDLINE=9903076; PubMed=9784498;
 RX Wells D.R., Tanguay R.L., Le H., Gallie D.R.;
 RA "HSP101 functions as a specific translational regulatory protein whose
 RT activity is regulated by nutrient status";
 RL Genes Dev. 12:3236-3251(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

Gallie D.R., Tanouay R.L., Wells D.R.;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF083344; AAC83689.2; -;
InterPro: IPR003593; AAA_ATPase.
InterPro: IPR003959; AAA_ATPase_central.
InterPro: IPR001270; Chaperin_c1pA/B.
InterPro: IPR004176; Clp_N.
Pfam: PF00004; AAA; 2.
Pfam: PF02861; Clp_N; 2.
PRINTS: PR00300; CLPPROTEASEA.
SMART: SM00382; AAA; 2.
PROSITE: PS00870; CLPAB_1; 1.
PROSITE: PS00871; CLPAB_2; 1.
ATP-binding; Heat shock.

SEQUENCE 918 AA; 101115 MW; 3450BF52A156E200 CRC64;
Query Match
Best Local Similarity 83.3%; Score 3790; DB 10; Length 918;
Matches 752; Conservative 86; Mismatches 70; Indels 12; Gaps 7;
1 MNPDKTHKNTETATTAHVLAVNAGHAQTPHLAAGALISDPTGIFPOAISSAGGENAA- 59
1 MNPDKTHKNTETATTAHVLAVNAGHAQTPHLAAGALISDPTGIFPOAISSAGGENAA- 59
60 -OSAEVINQALKLPSPSPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQIMGLL 118
61 GDSFVRLAGALKLPSPSPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQIMGLL 120
119 EDSQIRDLINVEGVATRVKSEVEKLGKE-GKKVESAGDTNFQALATYGRDLVEQAGK 177
121 EDAQADCLKEAGVSAVRALDKLGGDSNRKVESAGDTNFQALATYGRDLVEQAGK 180
178 LDPVIGRDEEIRVRVILSRRTKNNPVLIGPGVGTAVVEGLAQRIIVKGDVPSNLTDR 237
181 LDPVIGRDEEIRVRVILSRRTKNNPVLIGPGVGTAVVEGLAQRIIVKGDVPSNLTDR 240
238 LISLDMGALVAGAKYRGFEERLKSILKEVEDAEGKVLIFDEHLVLGAKTSGMDAA 297
241 LVALDMGALVAGAKYRGFEERLKSILKEVEDAEGKVLIFDEHLVLGAKTSGMDAA 300
298 NLFKPLARGQLRCIGATTLEBYRYKVEKDAFERRFOQVVAESPVDPTTILRLGKEK 357
301 NLFKPLARGQLRCIGATTLEBYRYKVEKDAFERRFOQVVAESPVDPTTILRLGKEK 360
358 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVQLDSQPBEIDNLE 417
361 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVQLDSQPBEIDNLE 420
418 KRMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKL 477
421 KRMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKL 480
478 KRMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKL 537
481 KRMQLELHALEREKDKASKARLIEVRKELDLRLKLOPLTMKYRKEKERIDEIRLKL 539
538 AEVVSRTGTIPVTRLGQNEKERLIGLADRLHKKRVVQGNQAVNVAEILASRAGLGRAQ 597
540 AEVVSRTGTIPVTRLGQNEKERLIGLADRLHKKRVVQGNQAVNVAEILASRAGLGRAQ 599
598 PTGSFLFGTGVGKTELAKALAEQLEDDENLVRIDMSYMEQHSVRLIGAPPGVGH 657
600 PTGSFLFGTGVGKTELAKALAEQLEDDENLVRIDMSYMEQHSVRLIGAPPGVGH 659
658 EGGQLTEAVRRPYCVILFDEVEKAHVAVENTLLQVLDGRLTDGQRTVDPRNVIIM 717
660 EGGQLTEAVRRPYCVILFDEVEKAHVAVENTLLQVLDGRLTDGQRTVDPRNVIIM 719
718 TSNLGAHLLAGTGVKTMVARDVNRVFRKFRPELLNRLDEIVVFDPLSHDQLKVA 777
720 TSNLGAHLLAGTGVKTMVARDVNRVFRKFRPELLNRLDEIVVFDPLSHDQLKVA 779
778 RLQMKDVAVRLAERGVAVLAVTDAALDYLLAESDVPYVGPARRRWEKKVVTLSKMWVR 837

Db 780 RLQMKDVAVRLAERGVAVLAVTDAALDYLLAESDVPYVGPARRRWEKKVVTLSKMWVR 839
QY 838 EIDENSTVYIDAGAGD-LVYRVV-SGGLVDASTGKSDVLIHANGP---KSDRAQA 891
Db 840 EIDENSTVYIDAGAGD-LVYRVV-SGGLVDASTGKSDVLIHANGP---KSDRAQA 891
QY 892 VKKWRIEIEIDDDNEEMIED 911
Db 900 VKKWRV--MEDGDEGMD 917

RESULT 9

Q9SY9 PRELIMINARY; PRT; 582 AA.
AC Q9SY9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 101 kDa heat shock protein (Fragment).
GN HSP101.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97134920; PubMed=8980480;
RA Boston R.S., Viitanen P.V., Vierling E.;
RT "Molecular chaperones and protein folding in plants.";
RL Plant Mol. Biol. 32:191-222(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99234099; PubMed=10216257;
RA Nieto-Sotelo J., Kannan K.B., Martinez L.M., Segal C.;
RT "Characterization of a maize heat-shock protein 101 gene, HSP101,
RT encoding a ClpB/Hsp100 protein homologue.";
RL Gene 230:187-195(1999).
DR EMBL: AF083327; AAD26530.1; -;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001270; Chaperin_c1pA/B.
DR PRINTS: PR00300; CLPPROTEASEA.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00871; CLPAB_2; 1.
KW ATP-binding; Heat shock.
FT NON_TER
SQ SEQUENCE 582 AA; 65767 MW; 9ABEAC1FAFE26B7C CRC64;

Query Match
Best Local Similarity 55.0%; Score 2502.5; DB 10; Length 582;
Matches 497; Conservative 49; Mismatches 32; Indels 5; Gaps 4;
QY 329 AFERRFOQVVAESPVDPTTILRLGKEKYECHGVRIQDRALINAAQLSARYITGRHLP 388
Db 1 AFERRFOQVVAESPVDPTTILRLGKEKYECHGVRIQDRALINAAQLSARYITGRHLP 60
QY 389 DKAILDVEACANVRVQLDSQPBEIDNLERKMQLELHALEREKDKASKARLIEVRKE 448
Db 61 DKAILDVEACANVRVQLDSQPBEIDNLERKMQLELHALEREKDKASKARLIEVRKE 120
QY 449 LDDLRDKLOPLTMKYRKEKERIDEIRLKKORRELMTFSLQAEERRYLARADLRGAIQ 508
Db 121 LDDLRDKLOPLTMKYRKEKERIDEIRLKKORRELMTFSLQAEERRYLARADLRGAIQ 180
QY 509 EVESIAOLEGTSSEENVMLTENGVPEHIAEVSRTGTIPVTRLGQNEKERLIGLADRLH 568
Db 181 EIDRAISKLE-SETGENMLTETVGPQLAEVSVSRWTGTIPVTRLGQNDKRLVGLADRLH 239
QY 569 KRVVQGNQAVNVAEILASRAGLGRAQQTGSFLFGTGVGKTELAKALAEQLEDDEN 628
Db 240 QRVVGQTEAVSAEAVLRSRAGLGRPQPTGSFLFGTGVGKTELAKALAEQLEDDEN 299

```

629 LLVRIDMSEYMEHSHVSRLLIGAPPVGVGHEGGQTEAVRRRPPYCVILFDEVEKAHVAVF 688
300 LLVRIDMSEYMEHSHVSRLLIGAPPVGVGHEGGQTEAVRRRPPYCVILFDEVEKAHVAVF 359
689 NTLLOVLDGRITDGGRTVDPRNSVIIMTSLGAHLLAGLTGKVTNEVARDVCMREVR 748
360 NTLLOVLDGRITDGGRTVDPRNTVIIMTSLGAHLLAGLVGKNSMKVARDLVMOEVR 419
749 KHFRPELNLRLDEIVVFDPLSHDQLRKVARLQMKDVAVRLAERGVALAVTDAALDYILAE 808
420 RHFRPELNLRLDEIVVFDPLSHDQLRKVARLQMKDVAVRLAERGVALAVTDAALDYILAE 479
809 SYDPVYGARPIRRWEKKVTVLSKVVREIDENSTVIYDAGAG--DLVYRVF-SGGLV 865
480 SYDPVYGARPIRRWEKKVTVLSKVVREIDENSTVIYDAGAG--DLVYRVF-SGGLV 539
866 DASTCKSDVLIHLIANGPKRDAQAQVKKRI-REIEDDDNEE 907
540 NAETGMSKSDIILQVPTSTRDAAQAVKXKIMBEDEGDMDER 582

RESULT 10
023323 PRELIMINARY; PRT; 668 AA.
01-JAN-1998 (TrEMBLrel. 05, Created)
01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
Heat shock protein-like protein (Heat shock protein like).
D13375W OR A74G14670.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
eurosids II; Brassicales; Brassicaceae; Arabidopsi.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
Bevan M., Stiekema W., Murphy G., Wambutt R., Pohl T., Terry N.,
Kreiss M., Kavanagh T., Entian K.D., Rieger M., James R.,
Puigdomenech P., Hatzopoulos P., Obermaier B., Duesterhoft A.,
Jones J., Palme K., Ansorge W., Delseny M., Bancroft I., Mewes H.W.,
Schueller C., Chalwatzi N.,
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
EU Arabidopsis sequencing project;
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
EMBL; Z97336; CAB46061.1; -.
EMBL; AL161539; CAB78509.1; -.
InterPro; IPR0031593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_cent.
InterPro; IPR001270; Chaperin_clpA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 1.
PRINTS; PR00300; CLPPROTEASA.
SMART; SM00382; AAA; 1.
ATP-binding; Heat shock.
SEQUENCE 668 AA; 74043 MW; SEEA7697FC4AC365 CRC64;

Query Match 53.4%; Score 2431; DB 10; Length 668;
Best Local Similarity 70.8%; Pred. No. 6.2e-111;
Matches 499; Conservative 76; Mismatches 90; Indels 40; Gaps 6;

1 MNPEKTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAG-GENAA 59
1 MNDLRFDPNVKILASARSHMSLSHGQVTPHLGVTLLSILTSVFRATISAGDGISA 60
60 QSAERVINALKLSQSPPPDDIPASSSLIKVIRRAQAQKSGRDTHLAVDQIMGLLE 119
61 QSVNVINQSLYK-----TKRNLGDTKGVAVLVLSILE 95
120 DSQIRDLNLNEGVATARKVEKVEKLRGKGVESASGDTNFQALKTYGRDLVQAGKLD 179

```

```

Db DSQISDVLRKEAGVVPKVKSEVEKLR-----GEVILFAKTYGTDLVQAGKLD 144
QY PVIGRDEIRRVRIILSRRTKNNPVLIGPVGKTAIVGGLAQRIKVGDPVNSLTDRVLI 239
Db PVIGRDEIRRVRIILSRRTKNNPVLIGPVGKTAIVGGLAQRIKVGDPVNSLTDRVLI 204
QY SLDMGALVAGAKYRGFEERLKSVLKEVEDAGKVLFIIDEIHLVLGAGKTBGSDAANL 299
Db SLEFGAMVAGTTLRGQFEERLKSVLKAVEAQKVVLFIDEIHLGAGKASGSDAANL 264
QY FKPMALARGOLRCIGATTLEEKYKVEKDAAPERRQQVYVAEPSPDPTIILRGLKEKYE 359
Db LKPMALARGOLRCIGATTLEEKYKVEKDAAPERRQQVYVAEPSPDPTIILRGLKEKYE 324
QY GHGVRIQDRALINAQLSARYITGRHLFPDKAIDLVDRACANVRVQLDSQPEIDNLRK 419
Db GHGVRIQDRALVLSAQLSERVITGRRLFPDKAIDLVDRACANVRVQLDSQPEIDNLRK 384
QY RMQLEIEHLALEREK-DKASKARLIEVRKELDDLRLKLOPLTKYRKEKERIDEIRRLKQ 478
Db VMQLEIEHLALEREK-DKASKARLIEVRKELDDLRLKLOPLTKYRKEKERIDEIRRLKQ 444
QY KREELMFSLOEAEERYDLARAADLYGAIQEVESAIAOLEGTSSEENVMLTENVGPEHIA 538
Db NREDDLMIALQEAERHDPKAAVLKYGAIQEVESAIAKLE-KSAKONVMLTETVGPENIA 503
QY EVVSWRTGIPVTRLQNEKERLIGLADRLKRVVQGNQAVNAVEAILRSRAGLGRQOP 598
Db EVVSWRTGIPVTRLQNEKERLIGLADRLKRVVQGNQAVNAVEAILRSRAGLGRQOP 563
QY TGSFLFLGPTGVKTELAKALAEQFDENLLVRIDMSYNEQHSVRLGAPPG-VYGH 657
Db SGSLFLGPTGVKTELAKALAEQFDENLLVRIDMSYNEQHSVRLGAPPG-VYGH 623
QY REGQLTEAVRRRPPYCVILFDEVEKAHVAVENTLLQVLDGRLTD 702
Db REGQLTEAVRRRPPYCVILFDEVEKAHVAVENTLLQVLDGRLTD 668

RESULT 11
QBDJ40 PRELIMINARY; PRT; 871 AA.
AC Q8DJ40;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ClpB protein.
GN TLR1389.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005373; BAC08941.1; -.
KW Complete proteome.
SQ SEQUENCE 871 AA; 98389 MW; D19D08E1A680F5FB CRC64;

Query Match 50.5%; Score 2299.5; DB 16; Length 871;
Best Local Similarity 53.1%; Pred. No. 2.4e-104;
Matches 456; Conservative 177; Mismatches 209; Indels 17; Gaps 9;

QY 2 NPEKTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAAQS 61

```

5 NPNQFTEKAWAATARTFDLAKAQHQNLSEHLMKSL--BQGLAQIFQKAG--CSVQR 61

62 AERVINOALKLPSQSPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLDES 121 121

62 IRDLTDFEIRGQPKIS--HPSGVVLGQSLDLLOBAERARQFGDEFISTEHLVLAFAQDD 120 120

122 Q1--RDILINEVGVATARVKSVVEKLGRKEGKVESASGDTNFQALKVYGRD--LVEAOAGKL 178 178

121 RFGKKLPQDILGLEKVLREAIQQIRGSQ--KVTDQNPGEKYAALEKYGRDLTLLAQGKL 178 178

179 DPVIGRDEIRRVVRIILSRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDVPNSLTDVRL 238 238

179 DPVIGRDEIRRVQIILSRRTKNNPVLIGEPGVGKTAAIBGLAQIRVADVPSLDRQL 238 238

239 ISLDMGALVAGAKYGEFEERLSKVLKEVEDAEGKVLFFIDEITHVLGAGKTEGSDMAAN 298 298

239 IALDMGALTAGKYGEFEERLKAUVKEVDSNGQILFIDEITHVVGAGATQGMADGN 298 298

299 LFKPMLARGOLRCIGATTLFEYKRYVEKDAAFRRRQOVVAPSPVDDTISIIRGLKERY 358 358

299 LLKPMALARGELCIGATTLDEYKRYTEKDAALERRRQOVVDPSPVEDTISIIRGLKERY 358 358

359 EGHGVPIQDRALINAAQLSARYITGRHLPDKAIDLVDEACANVRVOLDSPPEIDNLR 418 418

359 EIHHGVKISDTALVAAATLSARYISDRFLPKAIDLVDEAAAKLMEITSKPEELDEIR 418 418

419 KRMQLETELHALEREKDKASKARLIEVRKLELDLDRKLQPLTMKYKEKERDEIRLQK 478 478

419 KILQLEMERLSLOKETSAASDRLEKLELADLKEEQSLNQAQWAEKVEDIRLQSIKE 478 478

479 KRELMPSLQEAERRYDLARAADRLYGAIVEASAQLEGTSSEENV----MLTENVP 534 534

479 ETEKWNIEIQEAERNYDINRAAEKLYGKLTLELHKLEAEAKLREIQVGRSILRDEVTE 538 538

535 EHTAEVVSWRTGTPTVRLGONEKBERLIGIADRLHKRVGQGNQAVNSAAILRSRAGLR 594 594

539 ADIAETISKWTGTPVSKLVESEAKLHLEEEHLKRVWGQDEAVSAVAIAIKRSRAGLAD 598 598

595 AQQPTGSFLPLGTPGVKTELAKALAEOLFDDENLAVRIDMGEYMEQHSVRLIGAPGY 654 654

599 PNEPIASFPLGTPGVKTELAKALAAFPMDTEALVRLDMGEYMEKHAIVSRLIGAPGY 658 658

655 VGHREGQJTEAVRRRPYCVIILFDEVEKAHVAVFNTLLQVLDGRLTDQGRGRTVDPRNSV 714 714

659 VGYDEGGQJTEAIRRRPYAVLFDTEKAHPDVFNVFLQILDDGRTVDSQGRTVDFKNVI 718 718

715 IIMTSLNLGAHLL--AGLTGKVTMEVARQVMEVRKHPRELLNRLDEIVVDFPLSHDQ 772 772

719 IIMTSGISQYILVAGDSDRSYSEMTNR--VMEAMRAHPRPELNRVDFEIFFHSLRKDQ 776 776

773 LRKVARLQMDVAVRLAERGVALVATDAALDYLAESYDPVYGARPIRRWMEKKVVTLS 832 832

777 LRQIVLQVORLQRLSDRHITLSLTEXAIDFLAEVGYDPVYGARPLKRAIQOKLETPIA 836 836

833 KMVREEDENSTVIYDAG 851

837 KAILRGDFDGTILVDVG 855

JLT 12
228

O00828 PRELIMINARY; PRT; 867 AA.
 O00828;
 01-JUL-1997 (TREMBLrel. 04, Created)
 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 Heat shock protein 100.
 HSP100.
 Leishmania donovani.
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=5661;
 [1]
 SEQUENCE FROM N.A.

720 NLGAHLLAGLTGKVTMEVARDVNRVVRKHFRPELNLRLDEIVVDFPLSHDQRLKRVARL 779
 719 NLGAQYQNDMTSPKAVEAQTQVMEVGEVKKFPEFENRLDDIILFSLGLKEMTGIDL 778
 780 QMKDVAVRLAERGVALAVTDAAALDYVLAESVDPVYGARPIRRWMEKKVTVLSKVVUREE 839
 779 ITEELYGLKXQDSRVSLEAKVYVLESADFADMGARPLRRWVKVKNITTELGRMTISQE 838
 840 IDENSTVVI 848
 839 LSPNSTVKV 847

SUULT 13
 Q8YJ91 PRELIMINARY; PRT; 931 AA.
 Q8YJ91
 01-MAR-2002 (TrEMBLrel. 20, Created)
 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 ATP-dependent Clp protease, ATP-binding subunit CLPB.
 BMEI0195.
 Brucella melitensis.
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Brucellaceae; Brucella.
 NCBI_TaxID=29459;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=16M / ATCC 23456 / Biotype 1;
 MEDLINE=20020109; PubMed=11756688;
 Deivecchio V.G., Kaputal V., Redkar R.J., Patra G., Mujer C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 Haselkorn R., Kyrides N., Overbeek R.;
 "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis";
 Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
 EMBL; A8009461; ALU51377.1; -
 InterPro; IPR003593; AAA ATPase.
 InterPro; IPR003959; AAA ATPase centr.
 InterPro; IPR001270; Chaprinin-clpA/B.
 InterPro; IPR004176; Clp_N.
 Pfam; PF00004; AAA; 1.
 Pfam; PF02861; Clp_N; 2.
 PRINTS; PR00300; CLPPTASEA.
 SMART; SM00382; AAA; 2.
 PROSITE; PS00870; CLPAB_1; 1.
 PROSITE; PS00871; CLPAB_2; 1.
 Complete proteome.
 SEQUENCE 931 AA; 103217 MW; C227C61BC13AC076 CRC64;

Query Match 50.4%; Score 2294; DB 16; Length 931;
 Best Local Similarity 53.5%; Pred. No. 4.9e-104;
 Matches 471; Conservative 151; Mismatches 244; Indels 14; Gaps 7;

1 MNPEKTHKNETIATATHELAVNAGHAPTPLHAGALISDPTGIFPOAISSAGGNAQA 60
 58 MNIEKTVRVRGFIQSAQTFALSSGNOQFTPEHLKVLIDDDDEGLAASLVERAGGR--VG 115
 61 SAERVINOAKKLPSQSPDDIPASSSLIKVIRRAQAQKSRGDTLAVDQLMGLL-- 118
 116 DVMRGLOSALEKLPKVSNGNDQLYSQPLAKVFSLABELASKAGDSFVTVRLLTALAME 175
 119 EDSQTRDLINVEGVATARKVSEKLRGKGGKKBESAGDTNFQALKTYGRDLVE--QAG 176
 176 KSAKTSILSAAGVTPALNRVINDMR--KGRADSASAESNYDALKKYARDLTFEDARAG 233
 177 KLDPVIGRDEIRRVRLSRKNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSLTDV 236
 234 KLDPVIGRDEIRRTIQVLSRKNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSLTDV 293
 237 RLISLDMGALVAGAKYRGFEERLKSVLKEVEDAEGKVLIFDIHLVGLNGKTEGSDA 296

Db 294 QLMALDMGALVAGAKYRGFEERLKSVLKEVEDAEGKVLIFDIHLVGLNGKTEGSDA 353
 QY 297 ANLFKPLMARGLCIGATTLEERYKYVEKDAAFERRFOQVVAEPSPDITISILRGLKE 356
 Db 354 SNLLKPALARGELHCVGATTLLEERYKYVEKDAALARRFPQFVDEPTVEDTISILRGLKE 413
 QY 357 KYEGHHGVRIQDRALINAQLSARIYTORHLPDKAIDLVDACANVRVQLDSQPEIDNL 416
 Db 414 KYEQHHKVRVSDSALVAAAMLSNRYITDRFLPKAIDLVDAAASRLRMQVDSKPELDEI 473
 QY 417 ERKRMQLELHALEREKDKASKARLIEVRKELDDLRDLQPLTKMYKKEKERTDEIRRL 476
 Db 474 DERIMQLKIERALKVETDAASKORLQRIEKLSDLEESAEELTAKWQAEKOKLGLAIDL 533
 QY 477 KQKRELMFSIQEARRRYDLAARADRYGAIQEVESATAQLEGTSSSEN--VMLTENVGP 534
 Db 534 KQLEBEARNALAIARNGEPQKAGELAYGTIPQLEKQLADAE--SQENKGLSEETVTP 590
 QY 535 EHIAEVSRWTGIPVTRLGQNEKERLGLADRLHVRVVGQNAVNAVSEAILRSRAGLGR 594
 Db 591 DRVAQVISRWGTGIPVDRMLEGEREKLEMEDEIGKRVVGQGEAVQAISKAVRRARAGLQD 650
 QY 595 AQOPTGSLFLGPTGVGKTELAKALAEQLPDENLVRIDMSEYMEQHSVRLIGAPPGY 654
 Db 651 PNRPIGSPFLGPTGVGKTELTKALASFLFODDTAMVRIDMSEFMKHSVRLIGAPPGY 710
 QY 655 VGHGEGGQLTEAVRRPYCVILFDEVEKARHVAVENTLQLVLDGRLTDGQRTVDFRNSV 714
 Db 711 VGYEGGVLTTEAVRRPYQVILFDEIEKAPDPVFNVLQVLDGRLTDGQHTVDFRNTV 770
 QY 715 IIMTSNLGAHELLAGLTGKVTMEVARDVNRVVRKHFRPELNLRLDEIVVDFPLSHDQRL 774
 Db 771 IIMTSNLGAELYIV-NLGENDDVETVRDVMGVVRAASPRPEFLNRVDEIILFHLRREDMG 829
 QY 775 KVARLQMKDVAVRLAERGVALAVTDAAALDYVLAESVDPVYGARPIRRWMEKKVTVLSK 834
 Db 830 AIVDIQMRQLYLLSDRKITVLEDDAREWLANKGDPAYGARPLKRVIQKEVQDPLAER 889
 QY 835 VREIDENSTVIIDAGADLVVRVSGGLVDASTGKSD 874
 Db 890 ILLGDILGSLVKITAGSDRLNFRPISGAFAAEPEREDE 929

RESULT 14
 Q9AEM5
 ID Q9AEM5 PRELIMINARY; PRT; 874 AA.
 AC Q9AEM5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DB ClpB protein.
 GN CLPB OR B1864.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 CX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21172887; PubMed=11274130;
 RA Kkaza E., Teyssier J., Ouahrani-Bettache S., Liautard J.P.,
 RA Koehler S.;
 RT "Characterization of Brucella suis clpB and clpAB mutants and
 RL participation of the genes in stress responses";
 RN J. Bacteriol. 183:2677-2681 (2001).
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,

[illegible]

```

239 LDMGALIAGAKYRGEPEERLKA VLNVEKESDNKIILFIVEIHLVLGAGKSGDAMDANLL 298
301 KPMLARGQLRCTIGATTLEEVYKVKVEKDAAFERRFOQVYVAEPSVPTTISILRGLKEKVEG 360
299 KPMLARGDVRTIGATTLEEVYKVKVEKDAAFERRFOQVYVAEPSVPTTISILRGLKEKVEG 358
361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQIDSQSPHIDNLERKR 420
359 HGVQITDKAVVAAQLAGRYITNRFLPKAIDLIDEACANVRVTLSRPAEIDALERKK 418
421 MOLEIELHALEREKDKASKARLIEVKEKELDDLDKLOPLTMKYRKEKERIDEIRLKQKR 480
419 RQLEIEERALQDRDKASAKERLKA VAEIOKVEEKLGLPLAKYQBERGRIDELQATQAKL 478
481 EELMFSLQBAERRYDLARAADLRYGAIQVEGSAIAQL-EGTSGSEENVMLTENVGPEHIAE 539
479 DEKKVKLEAERMRDMETAADLKRYVIPILODRIRSLKEAIEKQKATMLQGTVTGTDIAT 538
540 VVSRWTGIPVTRLGONEKERLIGLADRLHRRVVGQNAVNVSRAILRSRAGLGRAOQPT 599
539 VVSRWTNIPVTKLSOTERERLLHLADQLHLRVKQDEAVSRVAEAILRSRAGLARSDRPT 598
600 GSFLFLGPTGVGKTELAKALAEOLFDDENLVRIDMSYMECHSYSLIGAPPGYVGHEE 659
599 GSFLFLGPTGVGKTELSKAVASELFDKATMVKLMDSEYMEQHSVARLIGTPPGYVGHEE 658
660 GGQLEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDDGRLTDGQRTVDFRNSVIMTS 719
659 GGQLEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDDGRLTDGQRTVDFRNSVIMTS 718
720 NLGAHELLAGLTKVMEVARDCVMEVRKHFPELLNRLDEIVVDFDPLSHQRLKRVARL 779
719 NLGAQYLQNMDTSPKPYEVAQVMEVGRKFFRPEFINRLDDIILFRSLGKEMTGII DL 778
780 QMKDVAVRLAERGVALLDAALDYILAESYDPVYGARPIRWMKKVVTLSKMWVREE 839
779 ITEELNRLXQDSIRVSLTEAKQVLESAPFADMGARPIRWMKKVVTLSKMWVREE 838
840 IDENSTVYI 848
839 LSPNSTVKV 847

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

Protein - protein search, using sw model

February 13, 2004, 01:28:20 ; Search time 45 Seconds
(without alignments)
856.560 Million cell updates/sec

US-09-812-350-17
Select score: 4550
Sequence: 1 MNPEKTHKNETIATAHEL.....VKMKRIEIEDDDNEEMIED 911

Index table: BLOSUM62
Gapop 10.0, Gapext 0.5

328717 seqs, 42310858 residues

328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194	48.2	857	US-08-887-534A-47	Sequence 47, Appl
2	2194	48.2	857	US-09-272-431-47	Sequence 47, Appl
3	2173.5	47.8	875	US-09-107-532A-4537	Sequence 4537, Ap
4	2140	47.0	863	US-09-328-352-6730	Sequence 6730, Ap
5	2133	46.9	859	US-09-199-637A-281	Sequence 281, Appl
6	2132	46.9	859	US-09-252-991A-21413	Sequence 21413, A
7	2108	46.3	871	US-09-134-001C-3979	Sequence 3979, Ap
8	1912	42.0	872	US-09-138-452A-163	Sequence 163, Appl
9	1839	40.4	908	US-08-249-380-2	Sequence 2, Appli
10	1830	40.2	823	US-09-134-001C-4081	Sequence 4081, Ap
11	1810.5	39.8	866	US-09-040-843-2	Sequence 2, Appli
12	1810.5	39.8	866	US-09-040-843-4	Sequence 2, Appli
13	1789.5	39.3	1044	US-09-252-991A-22493	Sequence 22493, A
14	1745.5	38.4	672	US-09-040-843-4	Sequence 4, Appli
15	1745.5	38.4	672	US-09-040-843-4	Sequence 4, Appli
16	1682.5	37.0	830	US-09-621-855-2	Sequence 5709, Ap
17	1604	35.3	845	US-09-107-532A-5709	Sequence 458, Appl
18	1450.5	31.9	558	US-09-198-452A-458	Sequence 80, Appl
19	1427.5	31.4	745	US-09-107-532A-5687	Sequence 5687, Ap
20	1396	30.7	879	US-09-252-991A-22604	Sequence 22604, A
21	1388.5	30.5	641	US-09-724-623-75	Sequence 76, Appl
22	1387	30.5	1049	US-09-252-991A-17298	Sequence 17298, A
23	1333.5	29.3	701	US-08-923-511-2	Sequence 2, Appli
24	1333.5	29.3	701	US-09-416-874A-2	Sequence 2, Appli
25	1327.5	29.2	767	US-09-252-991A-31198	Sequence 31198, A
26	1325.5	29.1	761	US-09-328-352-7492	Sequence 7492, Ap
27	1271	27.9	903	US-09-328-352-6479	Sequence 6479, Ap

Sequence 25547, A
Sequence 15, Appl
Sequence 31500, A
Sequence 25328, A
Sequence 28426, A
Sequence 14, Appl
Sequence 390, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 32259, A
Sequence 2, Appl
Sequence 3, Appl

US-09-252-991A-25547
US-09-218-363-15
US-09-252-991A-31500
US-09-252-991A-25328
US-09-252-991A-28426
US-09-218-363-14
US-08-858-207A-390
US-08-466-390-4
US-08-470-950-4
US-08-467-781-4
US-08-195-487-4
US-08-483-924-4
US-09-452-294-1
PCT-US93-06160-4
US-09-741-150-4
US-09-252-991A-32259
US-09-592-054-2
US-08-938-105-3

ALIGNMENTS

RESULT 1

US-08-887-534A-47
; Sequence 47, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:
; APPLICANT: Holder, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,534A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-887-534A-47

Query Match 48.2%; Score 2194; DB 4; Length 857;
Best Local Similarity 51.7%; Pred. No. 8e-175;
Matches 447; Conservative 154; Mismatches 239; Indels 24; Gaps 9;
Qy 1 MNPEKTHKNETIATAHELAVNAGHAQTPHLAGALISDPTGIPQAISSAGGENAAQ 60
Db 1 MRDLRTNKQLALADAQSLALGHNDQFTEPLHMGALLNQEGLSVPLTSA-GINAGQ 59
Qy 61 SAERVINQALKLPSPQPPDDTPASSLLIKVIRRAQAQKSRGDTHLAVDQIMGL 120

Fri Feb 13 10:59:40 2004

```

60 -LRTDINQALNLPQVEGTGGDVQSDQLVRVNLCDNVAQKRGDNFISSELFVLAALLES 118
121 -SQIRDLLNEVGAVATARKVSEVEKLRGEGKVESASGDTNFQALTKYGRDLVQEA--GK 177
119 RGTVDILKAAAGATTANITQATEQMRG--GESVNDQGAEDQORQALKKYTIDLTERAEQK 176
178 LDPVIGRDEIRVRVILSRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 237
177 LDPVIGRDEIRVIRTIQVLRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 236
238 LISLDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEHLVLGAGKTGSDMAA 297
237 VLALDMGALVAGAKYGEFEERLKGVLNDLAKQEGNVILFIDELHTMVGAGKADGAMDAG 296
298 NLKPKMLARGQLRCIGATTLEEVKRVKVEKDAFERRFQOVVVAEPSVPTTISILRLKEK 357
297 NMLKPALARGELHCVGATTIDEVYQVIEKDALEERFQKVFVAEPSVEDTIALRLKLER 356
358 YEGHHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEEIDNLE 417
357 YELHHVQITDPAIVAAATLSHRYIADRLQPKAIDLIDEAASSIRMQIDSKPEELDRUD 416
418 RKMQLIEIHLALEREKDKASKARLIEVRKELDDLRDKLQPLTMKVKERIDEIRLKL 477
417 RRIIQLKLEQQAALMKSDEASKKRLDMLNEELSDKEROYSELEENKAEKASLSGTQTIK 476
478 QKREELMFSIQEARRVYDLARADLYRGATQVE---SATAQLEGTSSEENVMLTENVG 534
477 AELEQAKIAIEQARRVGDLMARSELQYKIPLEKQLEAATQLEGKTM---LLRNKVT 533
535 EHIAEVVSRWTGIPVTRLQNGEKRLIGLADRLHVRVQNGQAVNAVSEAILRSRAGLGR 594
534 AEIAEVLARWTGIPVSRMSESEREKLRMEQLHRRVIGQNEAVDAVSNAIRSRLAGLAD 593
595 AQOPTGFLFGPTGVGKTELAKALBQLFDDENLVRIDMSEYMEQHSVSRILGAPPGY 654
594 PNRPIGSLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEFMKHSVSRILGAPPGY 653
655 VGHEGGQLTAVRRRYPYCVILFVEVEKAVAVENTLQVLDGRLTDGQRTVDPRNSV 714
654 VGYEGGYLTAVRRRYPYVILDEVEKAPDPVFNILLQVLDGRLTDGQRTVDPRNTV 713
715 LIMSNGABHLAAGLTKGVMEVARDVCMREVRKFRPELNLRLDRIVVPDPLSHDQLR 774
714 VIMTNSLGD-LIQERFELDYAHMKELVGVVSHNFRPEFNIRIDEVVFHPLGEQHIA 772
775 KVARLQMDVAVRLAERGVAVLTAADLYLAEVSDPYVYCARIRRMMEKVTLSKM 834
773 STAQIQLKRLYKLEERGVEIHISDEALKJLSENGDPVYGARPLKRAIOOOIENPLAQ 832
835 -----VVREIDENSTVYI 848
833 ILSGELVPGKVRIRVEDRIVAV 856

```

RESULT 2
 S-09-527-431-47
 Sequence 47, Application US/09527431
 Patent No. 6485899
 GENERAL INFORMATION:
 APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/527,431
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US/08/887,534
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 28341/33996
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: (312) 474-6600
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 857 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-527-431-47

Query Match 48.2%; Score 2194; DB 4; Length 857;
Best Local Similarity 51.7%; Pred. No. 8e-175;
Matches 447; Conservative 154; Mismatches 239; Indels 24; Gaps 9;

QY 1 MNPRFTHKTNETTATTAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSAGENAAQ 60
DB 1 MRLDLTKNFQLALADAQSLAGHDNQFIEPLHMSALLNQEGSVSPLLTSA-GINAGQ 59
QY 61 SAERVINQALKKLPSQSPDDIPASSSLIKVIRBAQAQKSRGDTHLAVDQIMGLLED 120
DB 60 -LRTDINQALNLPQVEGTGGDVQSDQLVRVNLCDNVAQKRGDNFISSELFVLAALLES 118
QY 121 -SQIRDLLNEVGAVATARKVSEVEKLRGEGKVESASGDTNFQALTKYGRDLVQEA--GK 177
DB 119 RGTVDILKAAAGATTANITQATEQMRG--GESVNDQGAEDQORQALKKYTIDLTERAEQK 176
QY 178 LDPVIGRDEIRVRVILSRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 237
DB 177 LDPVIGRDEIRVIRTIQVLRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 236
QY 238 LISLDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEHLVLGAGKTGSDMAA 297
DB 237 VLALDMGALVAGAKYGEFEERLKGVLNDLAKQEGNVILFIDELHTMVGAGKADGAMDAG 296
QY 298 NLKPKMLARGQLRCIGATTLEEVKRVKVEKDAFERRFQOVVVAEPSVPTTISILRLKEK 357
DB 297 NMLKPALARGELHCVGATTIDEVYQVIEKDALEERFQKVFVAEPSVEDTIALRLKLER 356
QY 358 YEGHHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEEIDNLE 417
DB 357 YELHHVQITDPAIVAAATLSHRYIADRLQPKAIDLIDEAASSIRMQIDSKPEELDRUD 416
QY 418 RKMQLIEIHLALEREKDKASKARLIEVRKELDDLRDKLQPLTMKVKERIDEIRLKL 477
DB 417 RRIIQLKLEQQAALMKSDEASKKRLDMLNEELSDKEROYSELEENKAEKASLSGTQTIK 476
QY 478 QKREELMFSIQEARRVYDLARADLYRGATQVE---SATAQLEGTSSEENVMLTENVG 534
DB 477 AELEQAKIAIEQARRVGDLMARSELQYKIPLEKQLEAATQLEGKTM---LLRNKVT 533
QY 535 EHIAEVVSRWTGIPVTRLQNGEKRLIGLADRLHVRVQNGQAVNAVSEAILRSRAGLGR 594
DB 534 AEIAEVLARWTGIPVSRMSESEREKLRMEQLHRRVIGQNEAVDAVSNAIRSRLAGLAD 593
QY 595 AQOPTGFLFGPTGVGKTELAKALBQLFDDENLVRIDMSEYMEQHSVSRILGAPPGY 654
DB 594 PNRPIGSLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEFMKHSVSRILGAPPGY 653

```


LENGTH: 863
TYPE: PRT
ORGANISM: Acinetobacter baumannii
09-328-352-6730
Query Match 47.0%; Score 2140; DB 4; Length 863;
Best Local Similarity 49.3%; Pred. No. 2.7e-170;
Matches 424; Conservative 181; Mismatches 245; Indels 10; Gaps 6;
1 MNPEKTHKNTIATATHELVAGHAQFTPLHAGALISDPTGIFPOAISSAGGENAAQ 60
5 MRFEKFTNRLOALSDAOLAMGKHDTATAGIHILSTLLEPSNI--SLLOQAGAR--LP 60
61 SAERVINOALKPLSOPSPDDIPASSSLIKVIRRAQAQSRGDTHLAVDQIMGLLE 120
61 ELKQKLEQALKDAPTITGDNVNLNPEAKALNADRYAQAGDEFISTDWILGLAET 120
121 SQIRDLLNEVGATARVSEKLEKGGKVESASGDTNFQALNTYGRDLVEQA--GKL 178
121 GETKNLSAVGTPDLSRKVIENIRGSD--KWSNNHEDQDRLNKYITIDLTERALSGKL 178
179 DPVIGRDEIRVRILSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGDVPSNLTVDRL 238
179 DPVIGRDEIRRTIQLSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGDVPSNLTVDRL 238
239 ISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHLVLAGKTEGSDAAN 298
239 LSLLDLSLAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHLVLAGKTEGSDAAN 298
299 LFKPMLARGOLRCIGATTILEYRKVEKDAAFERRPQQVYVAEPSVDPDTISILRGLKEKY 358
299 MLKPALARGELRCVGTATLDEYRQYTEKDAALERRFQKVLVDEPSVEDTIALRGLKEKY 358
359 EGHGVRIQDRALINAQOLSAVITGRHLPDKALIDVDEACANVRVQDLSQEPEDINLER 418
359 ATHGVQVILDSALIAAKKSHRYITDQLPDKALIDVDEACANVRVQDLSQEPEDINLER 418
419 KRMOLIEIHLEREXDKAKARLIEVRKELDLRDKLOPLMTYKKEKERIDEIRRLKQ 478
419 RLILQKMLQAEVKEDEKAGKAEVTHLEKQIAEVEYNDLEVRWKAETLVEGTQKQV 478
479 KREELMPSLOEARRVDLARAADLYRGAIOVESAJAQL--GTSSSENVMLENTVGPHEI 537
479 ELDKARTAPEKAREGLDAEALRQYVPELOKQLEQDEVAEENEEPKLIIRTKVTENEI 538
538 AEVSVRTGIPVTRLGONKBERLIGLADRLHVRVQGNQAVNAVSAIILSRAGLGRAQ 597
539 AEVVSATGIPVAKMMQGEREKULHMEEFVHDRVVGQDEAVAVNAVRSRAGLSDPNE 598
598 PTGSFLPLGPTGVGKTELAKALAEQFLDDENLVLRIDMSYMEQHSVRLIGAPPGYVGH 657
599 PSQSFLPLGPTGVGKTELAKALAEQFLDDENLVLRIDMSYMEQHSVRLIGAPPGYVGH 658
658 EGGQLFEAVRRPYPVILFDEVEKAVAVNFLLQVLDGRITDQGRVDPFRNSVILM 717
659 EGGVLTEAVRRKPYSVLVFDEVEKAPDPVFNILLQVLDGRITDQGRVDPFRNTVIM 718
718 TSNLGAELHLAGLTKGVTMVARDCVMEVRKHFPELNLRLDEIVVDFPLSHDQLRKVA 777
719 TSNLGSQVRELGEAGTDDV--RTVNVNVSQHFPEFNRIDELVIFSLKKAQIRGTA 777
778 RLQMKDVAVLAERGVAVLATAADLYLAESVDPVYGARPIRRWKEKKVYVTELSKNVVR 837
778 DIQLDLRLSRVLDRMSLTVDSDAFLLIDAGFPVYGARFLKRAIQQVVENTLAQKILS 837
838 BEIDENSTVIIDAGDLVY 857
838 GDFVAGDTILVKGNGHLVF 857

Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199,637A
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 281
LENGTH: 859
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-281

Query Match 46.9%; Score 2133; DB 4; Length 859;
Best Local Similarity 49.5%; Pred. No. 1e-169;
Matches 427; Conservative 176; Mismatches 245; Indels 14; Gaps 6;

QY 1 MNPEKTHKNTIATATHELVAGHAQFTPLHAGALISDPTGIFPOAISSAGGENAAQ 60
DB 6 MRIDLTSKQLALSDAOLAMGKHDTATAGIHILSTLLEPSNI--SLLOQAGAR--LP 65
QY 61 SAERVINOALKPLSOPSPDDIPASSSLIKVIRRAQAQSRGDTHLAVDQIMGLLE 119
DB 66 RSG--LNKELDALPKIQSPGTVNLSQDIARLLNQADRLAQKGDQFISSELVLAAMDE 123
QY 120 SQIRDLLNEVGATARVSEKLEKGGKVESASGDTNFQALNTYGRDLVEQA--GK 177
DB 124 NTRLGKLLGGVSRKALENANVLRG--GEAVNDPVVEESRQALDKYTVDMTKRAEKG 181
QY 178 LDPVIGRDEIRVRILSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGDVPSNLTVDVR 237
DB 182 LDPVIGRDEIRRTIQLSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGDVPSNLTVDVR 241
QY 238 LISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHLVLAGKTEGSDAAN 297
DB 242 LLALDMGALVAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHLVLAGKTEGSDAAN 301
QY 298 NLFPKMLARGOLRCIGATTILEYRKVEKDAAFERRPQQVYVAEPSVDPDTISILRGLKEK 357
DB 302 NMLKPALARGELRCVGTATLDEYRQYTEKDAALERRFQKVLVDEPSVEDTIALRGLKER 361
QY 358 YEGHGVRIQDRALINAQOLSAVITGRHLPDKALIDVDEACANVRVQDLSQEPEDINLE 417
DB 362 YEVHGVISITGALIAAKKSHRYITDQLPDKALIDVDEACANVRVQDLSQEPEDINLE 421
QY 418 KRMOLIEIHLEREXDKAKARLIEVRKELDLRDKLOPLMTYKKEKERIDEIRRLK 477
DB 422 RELIQLKIERALKKEDDEATRKELAKLEEDIVKLEREYADLEEIWESEKAEVQGSQAIQ 481
QY 478 KREELMPSLOEARRVDLARAADLYRGAIOVESAJAQL--GTSSSENVMLENTVGPHEI 537
DB 482 QKIEQAKQEMAEARRKGDLESARIQVTPDLRSIQMDVMDHOKTENQLLRNKVTOEEI 541
QY 538 AEVSVRTGIPVTRLGONKBERLIGLADRLHVRVQGNQAVNAVSAIILSRAGLGRAQ 597
DB 542 AEVSVRTGIPVTRLGONKBERLIGLADRLHVRVQGNQAVNAVSAIILSRAGLGRAQ 601
QY 598 PTGSFLPLGPTGVGKTELAKALAEQFLDDENLVLRIDMSYMEQHSVRLIGAPPGYVGH 657
DB 602 PSQSFLPLGPTGVGKTELAKALAEQFLDDENLVLRIDMSYMEQHSVRLIGAPPGYVGH 661

```

422 RRLIQLKIERBAKKEDEATRRKRLAKLEEDIVKLEREVADLEDIKWSEKAEVQGSQAIIQ 481
478 OKREELMFSLQEAERYDYLAARAADLFYGAIQEIVESAIAQLGTSSEENVMLTENVGPEHI 537
482 QKIEQAKQEMEAARRKGDLESWARIQYQITPDLERSLQWVDQHGKTENQLLRNKTDEEI 541
538 AEVVSRTWGTGIPVTRLGQNKERLIGLADRLHKRVGVGONQAVNAVSEAILRSRAGLGRAQQ 597
542 AEVVSQWGTGIPVSKMLGREGREKLIRMEQELHRRVIGODEAVAVNAVRSRAGLADPNR 601
598 PTGSFLFLGPTGVGKTELAKALAEQLFDDENLVRIDMSYMEQHSVSRSLIGAPPVGVGH 657
602 PSGSFLFLGPTGVGKTELCKALAEFLFDTREALVRIDMSSEFMKHSVARLIGAPPVGVGF 661
658 EREGQITEAVRRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGCRVDFRNSVLIIM 717
662 EREGGYLTEAIRRKPYSVWLLIDEVKAHPDVFNILLQVLEDGRLTDSHGRTVDFRNTVVVM 721
718 TSNLGAHL--LAGLTGKVTMEVARCVMEVEKHPPELLNRLDELTVVDPDLSHDQLRK 775
722 TSNLGAQIQEJAG-----DREAQRAAVMDAVNAHFRPEFINRIDEVVVEPLAREQIAG 776
776 VARLQKMDVAVRLAERGVALAVTDAALDYTLAESYDFVYGARPIRRMMEKKVVTLSKMV 835
777 IAEIQLGRLLKRLAEREISLELSQEALEDKLIAGVDFPVYGARPLKRAIQRWIENPLAQLI 836
836 VREEIDENSTVYIDAGADLVY 857
837 LAGKAPGASISAKVEGDEIVF 858

RESULT 7
US-09-134-001C-3979
; Sequence 3979, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3979
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3979

Query Match 46.3%; Score 2108; DB 4; Length 871;
Best Local Similarity 53.4%; Pred. No. 1.3e-167;
Matches 415; Conservative 167; Mismatches 173; Indels 22; Gaps 8;

Qy 86 SSSLIKVIRPAQAQKSRGTHLAVDQILMGLEDSDQIRDLINVEGVATARKVSEVEKLR 145
Db 88 SAKTNELLDAEKYMKSYEDFISMEHILRAADTDETTQ--KWVGKRVKVEIKSIITKVR 145
Qy 146 GKEGKKVESAGDTNFOALATYGRDIVEQA--GKLDPIVGRDEIRRVRLISRTKNKP 203
Db 146 G--GNHVTSGNPVEVYEALEKYGRLDVEFVQGNQDPVIGRDEIRNTILSRKTKNP 203
Qy 204 VLIGEPGVGKTAVVVEGLAQIRVKGDVPNSLTVDRLISLDMGALVAGAKYRGEFEERLKS 263
Db 204 VLIGEPGVGKTAIVVEGLAQIRVKDVPESLDDKTIPELDSLALVAGAKYRGEFEERLKAV 263
Qy 264 LKEVEDAGKVLIFIDBIHVLGAGKTEGSDAANLKPMLARGQLACIGATTLEEYRKY 323
Db 264 LXEYKSEGRGRIIFIDBIHMLVAGKGTGDAMDAGNMLKPMARGELHCIGATTLINEYREY 323

```

324 VEKDAAFERRFOVYVABSPVDTISILRLGLKEKYEHHGVRIODRALINAAQSAARYIT 383
324 IKDSALERRFOKVGVSFDPDENTISILRLGLKERVYVGVRIODRALVAAEASDRYIT 383
384 GHLPDKAIDLDEACANRVOLDSPQBEIDNLERKRMOLIEHLERKDKKASKALI 443
384 DRFLPDKAIDLVDQACATIRTEMGNSPTDELQVNRVVMQLEIESALKNESDNASKHLE 443
444 EYRKELDLRLDKLPMTKYRKEKERIDEIRLKKREELMPS---LOEARRRYDLARAA 500
444 ELQELSENEKEKOSLSKSVQEKK---IAVQEKRAELSSRQALDEAQTGNELEKAA 500
501 DLRYGAIQVEGSAIAQLE-----GTSSEENVMLTENUGPEHIAEVVSWTGPVTRLQON 555
501 ELQYGTIPQLEKELQEFEAQDEGTEDGSEMIREVVSDEIRGIDVSWQWGLPVSKLVET 560
556 EKERLIGLADRLHVRVQONAVNAVSEAILSRAGLGRAOQTSFPLGPTGVGKTEL 615
561 EREKLISLSDILHVRVQKAVDLVSDAVVRARAGIKDPNRPISGFLGPTGVGKTEL 620
616 AKALABQDFDENLVRIDMSYMEQHSVSRILGAPPGVYVGHGEGQLEAVRRPYPYVI 675
621 AKSLASLSDSKHMRIDMSYMEKHAVSRILGAPPGVYVGHGEGQLEAVRRPYPYVI 680
676 LFEVEKAHVAVFNTLLQVLDGRLTDGQRTVDPRNSVIIMTSLNGLAEHLGLTKYT 735
681 LLDVEKAHSDVFNVLQILDEGLTDSKGRSVDPKNTIIMTSLNIGSQVLENV--KDA 738
736 MEVARP---CVREVKHFRPELLNRLDEIVFDELSDHDLKRVARLQMKDVAVRLAERG 792
739 GEISDTEKAVMDSLHAYFKPILNRMDDIVLFKPLSVDDMSMIVDKILTQLNMLLDQH 798
793 VALAVTDAALDYILABSPVYVGPVGPVIRMEKVKVTELSKMWVREEDENSTVID 849
799 ISIEVTEAKWLGEEAYEPFGARPLKPVQRIETPIARMWIKESLPEGTLIKVD 855

USULT 8
-09-198-452A-163
Sequence 163, Application US/09198452A
Patent No. 655294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 163
LENGTH: 872
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
-09-198-452A-163

Query Match 42.0%; Score 1912; DB 4; Length 872;
Best Local Similarity 46.5%; Pred. No. 3.2e-151;
Matches 404; Conservative 171; Mismatches 271; Indels 22; Gaps 8;
4 EKPTHKNTETATIAHNAVAGHAQFPLHAGALISDPTGIFPQAISSAGGNAQAQSAE 63
8 EKFSDAVSEALEKAFELAKSGKTYVTNTHLLALLENTESFLYVLIKDHGNPG----- 62
64 RVINOALKKLPSQSP---PDDIPASSSLIKVIRRAQAQKSGDTHLAVDQILMGLLE 119
63 -LLNTAVKDALSRFTVVEGVDPKPSFGQLTLRDAKQAKTLGDEYISGDHLLAFWS 121
120 DSQ--IRDLLNEVGAVTARVKEVEKELKGGKVESASGDTNFOALKYGRDLVQQA-- 175
122 SNKEPFSWQTKTVSKFDLKNLTKIR--RGNRMDSPSABNSFQGLEKYCKNLALARE 179
176 GKLPDVIGRDEIRRVRIILSRRTKNNPVLIGERGCVKTAVEGLAQRIKVGDPNSLTD 235

180 GKLPDVIGRDEIRRTIQLSRRTKNNPMLIGEPGVGKTAIEGLAIRLIQGVDPESLKG 239
236 VRLISLDMGALVAGYRGFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTRGSM 295
240 KQIYVLDMGALVAGYRGFEERLKSVLKVESGGEHIFIDEVHTLVGAGATDGM 299
296 AANLFXPMARGQLRCIGATTLEBYKYVKEKDAAPERRQQVYVABSPVDPDTISILRLGLK 355
300 AANLLPALARGTLHCIGATTLEBYKYVKEKDAAPERRQQVYVABSPVDPDTISILRLGLR 359
356 EKYEGHGVRIODRALINAAQSAARYITGRHLDPDKAIDLDEACANRVOLDSPQBEIDN 415
360 EKYEIFHGVRIODRALINAAQSAARYITGRHLDPDKAIDLDEACANRVOLDSPQBEIDN 419
416 LERKRMOLIEHLERKDKKASKARLIEVRKELDDLRLDKLPMTKYRKEKERIDEIRR 475
420 KERELAALVYKQEAIRKEQSPSYQEEADAMQKSIDALRELASRLGWDBEKKLISGLKE 479
476 LKOKREELMPSLOEARRRYDLARAAADRLRYGAIQVEGSAIAQLEGT--SSEENVMLTENUGP 534
480 KONSLESMTSEBEARVADYNRAELRYSLIPQLEEEIKQDEASLNQRDLRLQBEVDE 539
535 EHTAAYVVRWTGIPVTELGQNEKERLIGLADRLHVRVQONAVNAVSEAILSRAGLGR 594
540 RLTAQVNVANWTGIPVQKMLEGEAKLILIEESLEERVVGQPPFAVSAVSDSIRAAVGLND 599
595 AQQPTGSGFLGPTGVGKTELAKALAEQDFDENLVRIDMSYMEQHSVSRILGAPPGY 654
600 PQRPLGVFLGPTGVGKTELAKALAEQDFDENLVRIDMSYMEQHSVSRILGAPPGY 659
655 VGHGEGQLEAVRRPYPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDPRNSV 714
660 VGYEGGSLSEALRRPYSVVLFDTEKADKEVINILLQVDFDDGILTDGKKRKNCKNAL 719
715 IIMTSLNGLAEHLGLTKYVMEVARDCVMREV---RKHFRPELLNRLDEIVVDFPLSH 770
720 FIMTSLNIGSP--LADYCKSGSELTKEAILSVSPVLYKRYLSPEFFMNRIDEILPFVPLTK 778
771 DOLRKVARLQMKDVAVRLAERGVALAVTDAALDYILABSPVYVGPVGPVIRMEKVKVTE 830
779 EDIVKIVGICQMRIRIAQLKARRINLSWDDSVILFSLQGYDSAFGARPLKRLIQCKVIL 838
831 LSKMWVREEDENSTVIDAGADLVYR 858
839 LSKALLKGDIKPDTSIETLWAKEVLVFK 866

RESULT 9
US-08-249-380-2
Sequence 2, Application US/08249380
Patent No. 5827685
GENERAL INFORMATION:
APPLICANT: Lindquist, Susan
TITLE OF INVENTION: Methods and Compositions of Genetic
TITLE OF INVENTION: Stress response Systems
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,380
FILING DATE:
CLASSIFICATION: 435

[illegible]

Sequence 2, Application US/09040843

GENERAL INFORMATION:

APPLICANT: Wang, Min

APPLICANT: Burnham, Martin

APPLICANT: Hodgson, John E.

APPLICANT: Rosenber, Martin
APPLICANT: Ward, Judith

TITLE OF INVENTION: MECH
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads

CITY: Philadelphia

COUNTRY: USA

COMPUTER READABLE FORM:

COMPUTER: IBM Compatible

SOFTWARE: FastSEQ for Windows

APPLICATION NUMBER: US/09/040
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/037,333
FILING DATE: 29-AUG-1997

ATTORNEY/AGENT INFORMATION;
NAME: Falk, Stephen T

REGISTRATION NUMBER: 36,752
REFERENCE/DOCKET NUMBER: GM10

TELEPHONE: 215-994-2488

TELEX:

825 KQ 826

ULT 12

09-621-855-2

Sequence 2, Application US/09621855

Patent No. 6346608

GENERAL INFORMATION:

APPLICANT: Jaworski, Deborah J.

Wang, Min

Shilling, Lisa K.

Burnham, Martin

Posberry, Andrew

Hodgson, John E.

Lawlor, Elizabeth

Rosenberg, Martin

Ward, Judith

TITLE OF INVENTION: MecB

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Dechert, Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2793

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/621,855

FILING DATE: 24-Jul-2000

CLASSIFICATION: <Unknown>

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 09/040,843

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Falk, Stephen T

REGISTRATION NUMBER: 36,795

REFERENCE/DOCKET NUMBER: GM10082

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488

TELEFAX: 215-994-2222

TELEX: <Unknown>

SEQUENCE CHARACTERISTICS:

LENGTH: 866 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

09-621-855-2

Query Match 39.8%; Score 1810.5; DB 4; Length 866;

Best Local Similarity 42.7%; Pred.No. 1e-142;

Matches 385; Conservative 164; Mismatches 252; Indels 101; Gaps 14;

5 KFTKNTETIATAHLAVNAGHAQTPPLHLAGALISDPTGIFPQAISAGGENAAQSAER 64

5 RUTERAQRVLAHQAEARLHNSNIGTBEHLGLLMKEPEGIAKVL-----ESFNITEDK 59

65 VINOAKKLPSQPPDDIPASSSLIKIRRAQAQKSGDTHLAVDQIMGLLEDQSQR 124

60 VIEVEKLIGHQDVGTHLYTPRAKKVIELSMDEARKLHNFVGTVHLGLIRENE-- 117

125 DLLNEVGVAAT---ARVKSVEKLKQKGGK-----KVESAGDTHNFQALKTVGRDL- 171

118 -----GVAARVFANLDLNIITKARARVVKALGNPMSNKAQAQSKSNNTPTLDSLRDLT 171

172 -VEQAGKLDLPVIGREEIRRVVRIILSRRTKNNPVLIGPGVGKTAVVGLAQRIVKGVDP 230

Db 172 VIAKGTGLDPVIGRDKETITRVIEVLSSRRRTKNNPVLIGEPGVGKTAIAEGLAQAVNVEVP 231

QY 231 NSLTDVRLISLDMGALVAGAKYRGEPEERLKSVLKEVEDAEGKVFILFIDIEHLVLGAGKT 290

Db 232 ETLKDKRVMSLDMGTVAGTKYRGEPEERLKKVMEIQQA-GNVILFIDELHTLVGAGGA 290

QY 291 EGSMDAANLFPKMLARGQLRCIGATTLEIYRYKVEKDAAPERRFQQVYVAVPSVPTTISI 350

Db 291 EGAIIDASNTLKPALARGELQCIGATTLDEYRKNIKDAALERRFPQVQVDEPSVVDVTVAI 350

QY 351 LRGLKEKYEGHGVRIODRALINAAQLSARYITGHLDPKADLDVDEACANVRVQLDSOP 410

Db 351 LKGLDRYEAHHRINISDEAIEAAVKLSNRYSDRFLPKAIDLIDIDEASSKVKLKSHTTP 410

QY 411 EEIDNLERKRMQLIEIHALEREKOKASKARLIEVRKLEDDLDLKLQPLTMKYRKEKERI 470

Db 411 NNLKEIEQ-----EIEKVKEKDAAVHAQFE---NNAANLRDKQTKLEKQY----- 453

QY 471 DEIRRLKQKRBEELMFSLOABRRYDLARAADLRYGAIQBVESAIAOLEGTSSSEENVMLTE 530

Db 454 -----EEAKNEW-----KNAQNGMST- 469

QY 531 NVGPEHIAEVVSRWTCIPVTRLGONEKERLIGLADRLHKRVVQONAVNAVSEAILRSRA 590

Db 470 SLSEEDIAEVIAWTGIPLTKNETSEKLLSLEDTLHERVIGQKQDANVSIKAVRARARA 529

QY 591 GLGAAQQTGSGFLGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVSRLLIGA 650

Db 530 GLKADPKRPIGSGFIFLGPTGVGKTELARALAESNFGDDAMIRVDMSEFMEKHAVALVGA 589

QY 651 PPGYVCHERGQQLTEAVRRPYCVILFDEVEKAHVAFTLLQVLDLDDGLTGGQGTVD 710

Db 590 PPGYVGHDDGGQLTEKVRKPKSVILFDEIEKAHPDVFNILLQVLDLDDGLTGGQGTVD 649

QY 711 RNSVLIWTSNLGAHL-----LAGLTCK---VTMEVARDCVMEVRKHPPELNLRLDEIV 763

Db 650 RNTIILMISNVGAQELQDQRFAGFGSSDGDQYETITKMLKELKNSFPPELNRVDDII 709

QY 764 VFDPLSHDQLRKVARLQMKDVAVRLAERGVALVATDAALDYTLAESYDPVYGARPIRRMM 823

Db 710 VFHKLTKELKEIVTMVWVKLTNRLSEQNINIIVTDKAKDKIAEGYDPEYGARPLIRAI 769

QY 824 EKKVTVTELSKAVVREEDENSIVYIDAGAGDLVYRVESGLVDASTGKKSVDVLIHANGP 883

Db 770 QKTIEDNLSELIDGNQIEGKVKTVVDHDKGFKY-----DIAEOTSETKTPSQAXILXNSP 824

QY 884 KR 885

Db 825 KQ 826

RESULT 13

US-09-252-991A-22493

; Sequence 22493, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 22493

; LENGTH: 1044

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22493

Query Match 39.3%; Score 1789.5; DB 4; Length 1044;
Best Local Similarity 43.8%; Pred. No. 7.6e-141; Indels 89; Gaps 19;
Matches 397; Conservative 168; Mismatches 252;

4 EKFTHTNETIATAHELAVNAGAOFTPLHLGALISDPTGIFPOAISAGGENAQAQSAE 63
199 EYFSKQTEYLQRAQVAEFGKREVDTEHLLYAL-ADADV--QAVLKQFGLSPAELKQ 255
64 RVINQAKKLPQSPDPDIPASSLSIKVIRRAQAQKSRGDTHLAVDOLINGL--LEDS 121
256 YIEANAVRG-ASKGEASEDMTTPSPVKSAQAFALSRGLSHYVGPHEHLLGLAAVDP 314
122 QIRDLNLEVGVATARKSVKRGKGGKVESASGDTNFQALKTGYGRDLVQEA--GKLD 179
315 FAGTLKKYGLTEQALQKAVKVGK-GAEDGRVDGSPNTPOLDKFSRDLTKLAREGKLD 373
180 PVIGRDEIRRVVRLSRRTKNNPVLIGEPGKTAIVEGLAQRIKVGDPVNSLTDVRLI 239
374 PVIGRSKEVETTIEVLARKKNNPVLIGEPGKTAIVEGLAQRIKVGDPVNSLTDVRLI 433
240 SLDMGALVAGAKYRGEFEERLKSVLKEVEDAGKVLFIIDEHLVGLGK--TEGSMDA 297
434 ELNINAVAGAKYRGEFEERLKSVLKEVEDAGKVLFIIDEHLVGLGK--TEGSMDA 493
298 NLFKPMIARGQLRCIGATTLEEVKRYKVEKDAFERRFOQVYVAEPVPTTISILGLK 357
494 NVLKPAMARGEMNLIGATTINVEQYKIEKDALERRFPQVFPVETVEGTISILGL 553
358 YEGHGVRIQDRALINAAQSAQRYITGRHLPKAIDLVDCAANVRVQLDSQPEIDNLE 417
554 LEGHKKVTIRDEAFVAAEELSDRYIGNRFLPKAIDLIDQAAARVIASTRPASTI 609
418 RKEMQLEIHLALEREKDAKARLIEVREKELDLKQPLTKMYRKEKERIDIRL 477
610 ---QELEALQKREQDYAASRKYDEAKVFE-----KKIQ 643
478 QKREELMFLQEAERRYDLARAADLYGAIQEVESAIAQLEGTSEENVMTENVGPHI 537
644 ERKEHL-----EQITERWQQTQSKTEE-----VRVEDI 672
538 AEVVSMTGIPVTRLGONKERLIGLADRLHVRVQGNQAVNAVEAILRSRAGIGRAQ 597
673 AEIISRLTGIPVTELTAEERELQMEERLHORVIGQEAITAVSDAVRLARAGLRQ 732
598 PIGSFLFGPTGKTELAKALAEQIFDENLVRIDMSEYMEQHSVSELIGAPGYVGH 657
733 PIATFLFGPTGKTELAKALAEVVGDEDMIRIDMSEYMERHVAVSRLIGAPGYV 792
658 EGGQLTEAVRRPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDPRNSVIIM 717
793 DEGGQLTEVRREPYSVILLDEIEKAHADVNILLQVFDGRLTDGKRVVDFNTIILA 852
718 TSNLGAHLL-----AGLTGKVTMEVARDVMEVRKHFRPELNLRLDEIVDPDLSH 773
853 TSNLSELIMKNAQAQGEFAQPEKLRKRE-LMTTLRGHFRPELNLRLDEIVDPESLS 911
774 RKVARKQMDVAVRLAERGVALAVTDAALDYILAESYDVPYGARPTRRMWEKKVVTLS 833
912 EDVRLQLEKRGKAAHAQDLYLHIDSLVGLHAEAYQPEFGARELKRQIRQOETRLAT 971
834 MYVREIDENSTV--YIDAGAGLVVRVESGGJLVADASTGKSPDLVTHANGPK-RSDAA 890
972 AMLKGEVKEGETVTFYDADKG-VGYR--KGAAPKPAARKKSGA---GETPKGRATAAR 1024
891 --AVKK 894
1025 KPAACK 1030

SULT 14

-09-040-843-4

Sequence 4, Application US/09040843

Patent No. 6124119

GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah J.
APPLICANT: Wang, Min
APPLICANT: Shilling, Lisa K.
APPLICANT: Burnham, Martin
APPLICANT: Fosberry, Andrew
APPLICANT: Hodgson, John E.
APPLICANT: Lawlor, Elizabeth
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: MecB
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,843
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/057,535
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10082
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 672 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-040-843-4

Query Match 38.4%; Score 1745.5; DB 3; Length 672;
Best Local Similarity 49.0%; Pred. No. 1.9e-137;
Matches 352; Conservative 129; Mismatches 166; Indels 71; Gaps 9;

QY 151 KVESASGDTNFQALKTGYGRDL--VEQAGKLDPVIGRDEEIRRVVRLSRRTKNNPVLIGE 208
DB 4 KNAQASKNNPTLDSLRDLTVAKDGTLDPVIGRDKETRVIEVLSRRTKNNPVLIGE 63
QY 209 PGVGTAVVEGLAQRIKVGDPVNSLTDVRLISLDMGALVAGAKYRGEFEERLKSVLKEVE 268
DB 64 PGVGTALAEGLAQIAVNVNEVPETLKDQVMSLDMGTVTVAGTKYRGEFEERLKSVMEEIQ 123
QY 269 DAEQKVLFIIDEHLVIGAGKTEGSDMAANLFRGMLARGQLRCIGATTILEEYKYVEKDA 328
DB 124 QA-GNVILFIDELTIVGAGAGCAIDASNILKPALARGELQICIGATTLLDEYRKNIEKDA 182
QY 329 AFERRFOQVYVABSPVPTTISILRLGKKEKYGHHGVRIQDRALINAAQSAQRYITGRHLP 388
DB 183 ALERRFOQVDEPESVVDTVAILKGLRDRYEAHHRINISDEATEAAVKLSNRVVSDFLP 242
QY 389 DKAILDVEACANVRVQLDSQPEIDNLERKRMQLTELHALEREKDKASKARLIEVRKE 448
DB 243 DKAILDIDEASSKVRKLSHTTNNLKEIEQ-----EIEKVKNEKDAVAHQBEF---N 292
QY 449 LDDLRDKLQPLTKMYRKEKERIDIRLKKRELMFSLQEAERRYDLARAADLYGAIQ 508

```

293 AANLRDKQTKLEKQY-----EAKNEW----- 314
509 EVESATIAQLEGTSSSENNVMTENVGPEHIAEVSRWTGIPVTRLGQNEKERLIGLADRLH 568
315 -----KNAQNGMST-SLSEEDIAEVAGTGIPLTKINETESEKLSLEDTLH 361
569 KRVVGQNAVAVSEAILRSRAGLGRAQOFTGSFLGPTGVGKTELAKALABQLFDDEN 628
362 ERVIGQKDAVNSISKAVRRARAGLKDPRIGSFIFLGPTGVGKTELAKALABQLFDDEN 421
629 LLVRIDMSYMEQHSVSRILGAPPGVGVGHEGGQLTAEVRRRPYCVILFDEVEKAHVAVF 688
422 AMIRVDMSEFMEKHAHSRLVAGPPGVGVGHDGQGLTEKVRKPYSVILFDETEKAHPDV 481
689 NTLLQVLDGRLTDGGRVDFRNSVIMTSLGAHLL-----LAGLTGK---VTMEVARD 741
482 NILLQVLDGRLTDGGRVDFRNSVIMTSLGAHLL-----LAGLTGK---VTMEVARD 741
742 CVMREVRKHFRPELNLRLDEIVVDFPLSHDQLRKVARLQMDQVAVRLAERGVALAVTDA 801
542 TMLKELKNSFRPEFLNRVDDIIVFHKLTKEELKEIVTMVNVKLITNRLSEQNINIIVTDA 601
802 LDYILAESYDPVYGARPIRMEKVKVVTLSKXVREIDENSTVIIDAGADLVYR 859
602 KDKIAEGYDPEYGARPLIRAIQKTIEDNLSELIDGNGQIEGKVTVDHDKGKEFYDI 659

```

JUL 15
 09-621-855-4
 Sequence 4, Application US/09621855
 Patent No. 6346608
 GENERAL INFORMATION:
 APPLICANT: Jaworski, Deborah J.
 Wang, Min
 Shilling, Lisa K.
 Burnham, Martin
 Fosberry, Andrew
 Hodgson, John E.
 Lawlor, Elizabeth
 Rosenberg, Martin
 Ward, Judith
 TITLE OF INVENTION: Mech
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESS: Dechert, Price & Rhoads
 STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103-2793
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/621,855
 FILING DATE: 24-Jul-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/040,843
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Falk, Stephen T
 REGISTRATION NUMBER: 36,795
 REFERENCE/DOCKET NUMBER: GM10082
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-994-2488
 TELEFAX: 215-994-2222
 TELEX: <unknown>
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 672 amino acids

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-621-855-4

```

```

Query Match
Best Local Similarity 38.4%; Score 1745.5; DB 4; Length 672;
Matches 352; Conservative 129; Mismatches 166; Indels 71; Gaps 9;

```

```

QY 151 KVESASGDTNFQALKTYGRDL--VEQAGKLDVPVIGRDEEIRRVVRIILSRRTKKNPVLIGE 208
DB 4 KNAQASKNNPTLTSLDRLDTVIADKDTLPVIGRDEEIRRVVRIILSRRTKKNPVLIGE 63
QY 209 PGVCKTAVVGLAQRIYKGVDPNSLTIVRLISLDMGALVAGAKYRGEPFERELKSVLKEVE 268
DB 64 PGVCKTAVVGLAQRIYKGVDPNSLTIVRLISLDMGALVAGAKYRGEPFERELKSVLKEVE 123
QY 269 DAEGKVLIFDEIHLVLGAGKTEGSDMAANLFKPLMARGQLRCIGATTLEERYKYVEKDA 328
DB 124 QA-GNVILFIDELHTLVGAGAGAGDAIDASNILPALARGELQICIGATTLDEYKNIKDA 182
QY 329 AFERRPOQVYVASEPVDITISILRLGKKEKTEGHGVRIODRALINAAQLSARYITGRHLP 388
DB 183 ALERRFQVQVDEPVSVDITVAILKGLRDRYEAAHRRINISDEATEAAVKLSNRYVSDRFLP 242
QY 389 DKAIDLVDACANVRVQLDSQPEIDNLERKRMQLLELHALEREKDKASKARLIEVRKE 448
DB 243 DKAIDLIDEASSKVRLLKSHPTPNLKEIEQ-----ETEKVKNKXDAVHAQEFN---N 292
QY 449 LDDLRDKLQPLTMKYRKEKERIDEIRRLKQREELMFSLQEAERYDLARAADRLRYGAIQ 508
DB 293 AANLRDKQTKLEKQY-----EAKNEW----- 314
QY 509 EVESATIAQLEGTSSSENNVMTENVGPEHIAEVSRWTGIPVTRLGQNEKERLIGLADRLH 568
DB 315 -----KNAQNGMST-SLSEEDIAEVAGTGIPLTKINETESEKLSLEDTLH 361
QY 569 KRVVGQNAVAVSEAILRSRAGLGRAQOFTGSFLGPTGVGKTELAKALABQLFDDEN 628
DB 362 ERVIGQKDAVNSISKAVRRARAGLKDPRIGSFIFLGPTGVGKTELAKALABQLFDDEN 421
QY 629 LLVRIDMSYMEQHSVSRILGAPPGVGVGHEGGQLTAEVRRRPYCVILFDEVEKAHVAVF 688
DB 422 AMIRVDMSEFMEKHAHSRLVAGPPGVGVGHDGQGLTEKVRKPYSVILFDEIEKAHPDV 481
QY 689 NTLLQVLDGRLTDGGRVDFRNSVIMTSLGAHLL-----LAGLTGK---VTMEVARD 741
DB 482 NILLQVLDGRLTDGGRVDFRNSVIMTSLGAHLL-----LAGLTGK---VTMEVARD 741
QY 742 CVMREVRKHFRPELNLRLDEIVVDFPLSHDQLRKVARLQMDQVAVRLAERGVALAVTDA 801
DB 542 TMLKELKNSFRPEFLNRVDDIIVFHKLTKEELKEIVTMVNVKLITNRLSEQNINIIVTDA 601
QY 802 LDYILAESYDPVYGARPIRMEKVKVVTLSKXVREIDENSTVIIDAGADLVYR 859
DB 602 KDKIAEGYDPEYGARPLIRAIQKTIEDNLSELIDGNGQIEGKVTVDHDKGKEFYDI 659

```

Search completed: February 13, 2004, 01:35:12
 Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Protein - protein search, using sw model

Search time 76 Seconds
(without alignments)
2509.832 Million cell updates/sec

US-09-812-350-17

ect score: 4550

1 MNPEFTHKTNETATATHEL.....VKMKRIEIEDDDNEEMIED 911

BLOSUM62

Gapop 10.0 , Gapext 0.5

801455 seqs, 209382283 residues

1 number of hits satisfying chosen parameters: 801455

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Alt	%	Score	Query Match	Length	ID	Description
1	3883.5	85.4	912	12	US-10-310-154-481	Sequence 481, App
2	2315	50.9	881	12	US-10-369-493-20047	Sequence 20047, A
3	2288	50.3	872	12	US-10-310-154-484	Sequence 484, App
4	2288	50.3	872	12	US-10-369-493-2762	Sequence 2762, App
5	2288	50.3	872	14	US-10-047-260-38	Sequence 38, Appl
6	2282	50.2	879	15	US-10-156-761-14770	Sequence 14770, A
7	2264.5	49.8	857	12	US-10-369-493-20488	Sequence 20488, A
8	2247	49.4	862	12	US-10-369-493-19762	Sequence 19762, A
9	2243	49.3	862	12	US-10-369-493-20953	Sequence 20953, A
10	2234.5	49.1	905	12	US-10-369-493-22555	Sequence 22555, A
11	2225	48.9	868	12	US-10-369-493-11982	Sequence 11982, A
12	2224	48.9	859	12	US-10-369-493-16834	Sequence 16834, A
13	2214.5	48.7	907	12	US-10-369-493-3990	Sequence 3990, App
14	2204	48.4	861	9	US-09-815-242-14088	Sequence 14088, A
15	2199.5	48.3	858	12	US-10-369-493-10608	Sequence 10608, A

Sequence 11393, A
Sequence 7831, App
Sequence 47, Appl
Sequence 483, App
Sequence 10258, A
Sequence 23508, A
Sequence 17826, A
Sequence 2769, App
Sequence 15599, A
Sequence 15976, A
Sequence 16333, A
Sequence 11134, A
Sequence 485, App
Sequence 17514, A
Sequence 8096, App
Sequence 9166, App
Sequence 9374, App
Sequence 6999, App
Sequence 5230, App
Sequence 12526, A
Sequence 19403, A
Sequence 19111, A
Sequence 5192, App
Sequence 19625, A
Sequence 281, App
Sequence 20217, A
Sequence 12048, A
Sequence 558, App
Sequence 13666, A
Sequence 18448, A

ALIGNMENTS

RESULT 1

US-10-310-154-481
Sequence 481, Application US/10310154
Publication No. US20030233670A1

GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D
APPLICANT: Chomet, Paul S.

APPLICANT: Adams, Thomas H
APPLICANT: Ruff, Thomas G.

APPLICANT: Agarwal, Ameeta K.
APPLICANT: Ahrens, Jeffrey E.

APPLICANT: Hall, James A.
APPLICANT: Bann, G.

APPLICANT: Bell, Erin
APPLICANT: Boddupalli, Raghava

APPLICANT: Deikman, Jill
APPLICANT: Deng, Molian

APPLICANT: Dong, Jinzhao
APPLICANT: Duff, Stephen M.

APPLICANT: Galligan, Meghan M.
APPLICANT: Hinchey, Brenda S.

APPLICANT: Huang, Shihshieh
APPLICANT: Johnson, G. Richard

APPLICANT: Jung, Vincent
APPLICANT: Kretzmer, Keith A

APPLICANT: Laccetti, Lucille B.
APPLICANT: Lai, Chao-Qiang

APPLICANT: Lee, Gary
APPLICANT: Lin, Jie-Yi

APPLICANT: Liu, Jingdong
APPLICANT: Lu, Bin

APPLICANT: Lund, Adrian
APPLICANT: Lund, Michael M.

APPLICANT: Madson, Linda L.
APPLICANT: Malloy, Kathleen A.

APPLICANT: McKiel, Christine L.
APPLICANT: Miller, Philip W.

APPLICANT: Padmavathi, Manchikanti

APPLICANT: Parnell, Laurence D.
 APPLICANT: Start, William G.
 APPLICANT: Tennessee, Dan
 APPLICANT: Vidya, K.R.
 APPLICANT: Wang, Haiyun
 APPLICANT: Xin, Zhanguo
 APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping
 APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li

TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants

FILE REFERENCE: 38-15(52796)B

CURRENT APPLICATION NUMBER: US/10/310,154

CURRENT FILING DATE: 2002-12-04

PRIOR APPLICATION NUMBER: 60/337,358

PRIOR FILING DATE: 2001-12-04

NUMBER OF SEQ ID NOS: 736

SEQ ID NO 481

LENGTH: 912

TYPE: PRT

ORGANISM: Zea mays

US-10-310-154-481

Query Match 85.4%; Score 3883.5; DB 12; Length 912;

Best Local Similarity 84.6%; Pred. No. 2.7e-268;

Matches 772; Conservative 76; Mismatches 58; Indels 7; Gaps 6;

1 MNPEKTHKTNETTATAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSA-GGENAA 59

1 MNPNDFTHKTNEAIVGAHEIAVEAGHAQFTPLHAAVLAADKGGILRAITGASGGDGA 60

60 -QSARVINOALKKLPSPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQILMGLL 118

61 GDSFVUNLSLKLPSPPDVPASTALIKVIRRAQAQKSGDTHLAVDQILMGLL 120

119 EDSQIRDLNNEVGATARKVESEKLRGEGKGVESASGDTNFQALTYGRDLVEQAGKL 178

121 EDSQISDCLKEAGVSAARVRAELEKLRGEGRRVESASGDTNFQALTYGRDLVEQAGKL 180

179 DPVIGRDEIRRVRLSRRTKNNPVLIGEGVGKTAIVEGLAQRIKGVDPNSLTDLVRL 238

181 DPVIGRDEIRRVRLSRRTKNNPVLIGEGVGKTAIVEGLAQRIKGVDPNSLTDLVRL 240

239 ISLDMGALVAGAKRGGEFEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTEGSDAAN 298

241 IALDMGALVAGAKRGGEFEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTEGSDAAN 300

239 LFKPLMARGQLRCIGATTLEBYRYKVEKDAAFERRFQQVVAEPSPVDTTISILRGLKEKY 358

301 LFKPLMARGQLRCIGATTLEBYRYKVEKDAAFERRFQQVVAEPSPVDTTISILRGLKEKY 360

359 EGHGVRTQDRALINAAGLSARYITGRHLPKADLDLVEACANRVQLDSOPEIDNLER 418

361 EGHGVRTQDRALVVAAGLSARYIMGRHLPKADLDLVEACANRVQLDSOPEIDNLER 420

419 KRMQLEIHELALEREKDKASKARLIEVKEKEDLDLKLQPLTMKYRKEKERIDIRLQK 478

421 KRIQLEIHELALEREKDKASKARLIEVKEKEDLDLKLQPLTMKYRKEKERIDIRLQK 480

479 KREELMFSLOEABRYDARAADLYGAIOVESAIQLETSSEENVMLTENVPPIHIA 538

481 RREELQFTLOEABRYDARAADLYGAIOVESAIQLETSSEENVMLTENVPPIHIA 539

539 EVVSRWTGTPVTRLQONKERLIGLADRLHKKVVCQNCANVAVSEAILRSAGLGRQOP 598

540 EVVSRWTGTPVTRLQONKERLIGLADRLHKKVVCQNCANVAVSEAILRSAGLGRQOP 599

599 TGSFLLFGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVRLIAGPPYVGHE 658

600 TGSFLLFGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVRLIAGPPYVGHE 659

QY 659 EGGQLTEAVRRRPYCVILFDEVEKARVAVFNTLLQVLDGRLTDGQRTVDPRNSVILMT 718
 DB 660 EGGQLTEAVRRRPYCVILFDEVEKARVAVFNTLLQVLDGRLTDGQRTVDPRNTVILMT 719
 QY 719 SNLGAHLLAGLTGKVTMEVARDVWREVRKHFRELLNRLDEIVFDPPLSHDQKRVAR 778
 DB 720 SNLGAHLLAGLVGKNSMKVARDLVMOEVRHFRPELLNRLDEIVFDPPLSHDQKRVAR 779
 QY 779 LQMKOVAVLAERGVALAVTDAALDYVLAESVDPVYGARPIRRWMEKKVVTLSKMWVRE 838
 DB 780 LQMKOVAVLAERGVALAVTDAALDYVLAESVDPVYGARPIRRWMEKKVVTLSKMWVRE 839
 QY 839 EIDENSTVYIDAGAG--DIVRYVE--SGGLVDASTGKSDVLIHIANGPKRSDAAQAVKQM 895
 DB 840 EIDENCTVYIDAAPGKDELIVYRDSGGLVNAETGMSDILIOVENSSTRSDAAQAVKQM 899
 QY 896 RI-EEIEDDDNEE 907
 DB 900 RIMEDEDDGDEE 912

RESULT 2

US-10-369-493-20047

; Sequence 20047, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 20047

; LENGTH: 881

; TYPE: PRT

; ORGANISM: No. US20030233675A1loc punctiforme

; US-10-369-493-20047

Query Match 50.9%; Score 2315; DB 12; Length 881;
 Best Local Similarity 53.9%; Pred. No. 2.1e-156;
 Matches 467; Conservative 172; Mismatches 199; Indels 28; Gaps 11;

QY 2 NPEKFTHTKTNETTATAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSAGGENAAQS 61

DB 12 NPEQTEKAWALVTRPTAKQFOHQIESEHLMALL-EQEGE-ASSIFNKAGVNVQKL 69

QY 62 ABRVINOALKKLPSQSPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQILMGLLSDS 121

DB 70 HERTIDFIRNPQKVGSSGSGVYIGHSERLDRAPQYRKEFGDEYIISIEHLILAFADK 129

QY 122 QI-RDLLNNEVGATARKVESEKLRGEGKGVESASGDTNFQALTYGRDLVEQA--GKL 178

DB 130 RFGKGLFQEGDLEKLRNIQQIRGSG--KVTQNPVKVEALEKYGREDLTOLAHEGIL 187

QY 179 DPVIGRDEIRRVRLSRRTKNNPVLIGEGVGKTAIVEGLAQRIKGVDPNSL----- 233

DB 188 DPVIGRDEIRRTIQLSRRTKNNPVLIGEGVGKTAIVEGLAQRIKGVDPNSL----- 247

QY 234 --TDVRLISLDMGALVAGAKRGGEFEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTE 291

DB 248 GRHRKLLIADMALTAGAKRGGEFEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTE 307

QY 292 GSDMAANLFPKMLARGQLRCIGATTLEBYRYKVEKDAAFERRFQQVVAEPSPVDTTISIL 351

DB 308 GSDMAANLFPKMLARGQLRCIGATTLEBYRYKVEKDAAFERRFQQVVAEPSPVDTTISIL 367

352 RGLKKEKYGHHGVIRIQDRALINAAQISARYITGRHLPKDAIDLVDACANVRVQDSQPE 411
 368 RGLKERYELHGHGVKISDSALVAATLSARYISDRFLPKDAIDLVDAAAALKEWITSKE 427
 412 EIDNLERKMOLEIEHLERKDKASKARLIEVRKELDDLDKQLPQTMKYRKEKERID 471
 428 ELDEIDRKILQLEMERLSQKETSASRELERELERELAEKQERQALNAQOAKGLIID 487
 472 EIRLKQKEELMFSIQEAERRYDLARAADRYG-----AIQEVESAIAQLEGTSEE 524
 488 RIRQIRQETERNVVEIQEAERDYDLNRAELKYSKLTQRLKEAEARLAQIQ-TSGKS 546
 525 NVMLTENVCPEHIAEVVSWTGIPTVRLQNEKERLIGLADRLHRVVGQNAVNSEA 584
 547 --LLREEVTEADIAETISKWTGIPVSRVSEMOKLHLEELHVRVIGQDEAVRAVADA 604
 585 ILRSRAGLGRAOQPGTSFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQSHV 644
 605 IQRSRAGLADPNRPASFLGPTGVGKTELAKALAEYLFDEDAVLRIDMSEYMEKHAV 664
 645 SELGAPPGVYEGGGQTEAVRRPYPYCVILFDEVEKAHVAVFNTLLQVLDGGRITDQG 704
 665 ARLGAPPGVYEGGGQTEAVRRPYPYCVILFDEVEKAHVAVFNTLLQVLDGGRITDQG 724
 705 GTVDPRNSVIITMNLGAHLL--AGLTGKVTMEVARDVWRVVRKHPRELINRLDEI 762
 725 GTVDPRNSVIITMNLGAHLL--AGLTGKVTMEVARDVWRVVRKHPRELINRLDEI 782
 763 VVFDPLSHDLQKRVARLQWQVAVLAERGVAVALTDAALDYLAESYDVPVGPAPIRRW 822
 783 IIFHSRKNDELREIVKLVQVORLEERLEERKLSKISDQALDWIVQVGVDPVYGPAPLRA 842
 823 MEKKVVELSKMVRREIDENSTVYI 848
 843 IQRELETPIAKILRGEEHGTIVV 868

LT 3
 0-310-154-484
 Sequence 484, Application US/10310154
 Publication No. US20030233670A1

GENERAL INFORMATION:
 APPLICANT: Edgerton, Michael D
 APPLICANT: Chomet, Paul S.
 APPLICANT: Adams, Thomas H
 APPLICANT: Ruff, Thomas G.
 APPLICANT: Agarwal, Ameeta K.
 APPLICANT: Ahrens, Jeffrey E.
 APPLICANT: Ball, James A.
 APPLICANT: Banu, G.
 APPLICANT: Bell, Erin
 APPLICANT: Boddupalli, Raghava
 APPLICANT: Deikman, Jill
 APPLICANT: Deng, Mollan
 APPLICANT: Dong, Jinzhao
 APPLICANT: Duff, Stephen M.
 APPLICANT: Galligan, Meghan M.
 APPLICANT: Hinchey, Brenda S.
 APPLICANT: Huang, Shihshieh
 APPLICANT: Johnson, G. Richard
 APPLICANT: Jung, Vincent
 APPLICANT: Kretzmer, Keith A.
 APPLICANT: Laccetti, Lucille B.
 APPLICANT: Lai, Chao-Qiang
 APPLICANT: Lee, Gary
 APPLICANT: Lin, Jie-Yi
 APPLICANT: Liu, Jingdong
 APPLICANT: Lu, Bin
 APPLICANT: Luehly, Michael M.
 APPLICANT: Lund, Adrian
 APPLICANT: Madson, Linda L.
 APPLICANT: Malloy, Kathleen A.
 APPLICANT: McKiel, Christine L.

APPLICANT: Miller, Philip W.
 APPLICANT: Padnavathi, Manchikanti
 APPLICANT: Parnell, Laurence D.
 APPLICANT: Start, William G.
 APPLICANT: Tennesen, Dan
 APPLICANT: Vidya, K.R.
 APPLICANT: Wang, Haiyun
 APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping
 APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li
 TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
 FILE REFERENCE: 38-15(52796)B
 CURRENT APPLICATION NUMBER: US/10/310,154
 CURRENT FILING DATE: 2002-12-04
 PRIOR APPLICATION NUMBER: 60/337,358
 PRIOR FILING DATE: 2001-12-04
 NUMBER OF SEQ ID NOS: 736
 SEQ ID NO 484
 LENGTH: 872
 TYPE: PRT
 ORGANISM: Synechocystis PCC6803
 US-10-310-154-484

Query Match 50.3%; Score 2288; DB 12; Length 872;
 Best Local Similarity 52.6%; Pred. No. 1.7e-154;
 Matches 460; Conservative 167; Mismatches 229; Indels 18; Gaps 7;

Qy 2 NPKFTHKTNETIATAHELAVNAGHAQFTPLHAGALISDPTGIPFOAISSAGGENAOS 61
 Db 5 DPNKFTKAWELAKTPEIAKQHRQOQIETHELLSALL--EONGLATSIENKACA-----S 58
 Qy 62 AERV---INQALKPLSQSPDPDIPASSSLIKVIRRAQAAQKSRGDTHLAVQDLMGLL 118
 Db 59 IPRVNDQVNSFIAQPKLGNPSEIYLGRLSLDKLLDNAETAKSKYGGDYISIEHLMAAYG 118
 Qy 119 EDSQI--RDILNEVGAVATARKVSEVKLRGKGGKVESAGDTNFQALKTGYGRDLVQEA-- 175
 Db 119 QDRGLGNLYREIGLTENKLAELIKIRGTQ--KVTDQNPGEKYSELEKYGRDLTELARE 176
 Qy 176 GKLDPIVGRDEIRRVVIRLSRRTKKNPVLIGEPGKTAIVVEGLAQRIIVKGVDPNSLTD 235
 Db 177 GKLDPIVGRDEIRVIRTIQLSRRTKKNPVLIGEPGKTAIAEGLAQRIINHDPESLSD 236
 Qy 236 VRLISLDMGALVAGKVGCEFFERLKSVLKEVEDAEKGVILFIDEIHLVIGAKTEGSD 295
 Db 237 RKLLISLDMGALVAGKVGCEFFERLKSVLKEVEDAEKGVILFIDEIHLVIGAKTEGSD 296
 Qy 296 AANLFPMLARGQLRCIGATTLEERYKVEKDAAFERRPQVYVAPSPVDPDTISILRGLK 355
 Db 297 AGNLLFPMLARGALRCIGATTLEERYKVEKDAAFERRPQVYVAPSPVDPDTISILRGLK 356
 Qy 356 EYEGHHGVIRIQDRALINAAQISARYITGRHLPKDAIDLVDACANVRVQDSQPEEIDN 415
 Db 357 ERYEVHGVKVIADSLVAAMLSNRYISDRFDPKDAIDLVDAAAALKEWITSKEPELDE 416
 Qy 416 LERKRMOLRIELHALEREKDKASKARLIEVRKELDDLDKQLPQTMKYRKEKERIDETR 475
 Db 417 VDRKILQLEMERLSQKETSASRELERELERELAEKQERQALNAQOAKGLIIDQIRT 476
 Qy 476 LKQREELMFSIQEAERRYDLARAADRYGATQEVESAIAQLEGTSSSEENV---MLTEN 531
 Db 477 VKETIQVNLIEIQOQORDYDNKAAELQVGLTDLQORVLEALETQLAEQOTSGKSLIRE 536
 Qy 532 VGPEHIAEVVSWTGIPTVRLQNEKERLIGLADRLHRVVGQNAVNSEAAILERSRAG 591
 Db 537 VLESIDIAETISKWTGIPISKLVESEKELHLEELHVRVIGQDEAVRAVAEAIQSRAG 596
 Qy 592 LGRAOQPGTSFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQSHVSLIGAP 651

597 LSDPNRPTASFIPLGPTGVGKTELAKALAKNLFDTTEALVRIDMSEYMEKHAVALRMGAP 656
652 PGYVHEGGQTEAVRRPYPYCVILFDEVEKARHVAFTLLQVLDGRLTDGQRTVDFR 711
657 PGYVHEGGQTEALRRPYPYVILFDEVEKAGHGVNMLQILDDGRLTDGQRTVDFR 716
712 NSVIMTSLGAEHLIAGLTKVMEVARDCVNRVRRKHFPELNLRLDEIVVDFPLSHD 771
717 NTIIMTSLGQYILDVAGDDSRVEEMRSRVMDVNRNFRPEFLNRVDETIIFHGLQKS 776
772 QLRKVARLQMKDVAVRLAERGVVALVTDAAALDYILAESYDVPYVGPPIRMMWKKVTEL 831
777 ELRSIVQIQSLATLEBQKLTLDKALDFLAAGVDPYVGPPIRMMWKKVTEL 836
832 SKOVVREIDENSTVYIDAGADLVYRVESGGIV 865
837 AKGILRGDYKPGETIVVDETDRLSFTSLRGDLV 870

RESULT 4
10-369-493-2762
Sequence 2762, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 2762
LENGTH: 872
TYPE: PRT
ORGANISM: Synechocystis sp.
10-369-493-2762

Query Match 50.3%; Score 2288; DB 12; Length 872;
Best Local Similarity 52.6%; Pred. No. 1.7e-154;
Matches 460; Conservative 167; Mismatches 229; Indels 18; Gaps 7;
2 NPEKTHKTNETIATAHVAHAGHQAFTPLHAGALISDPTGIFPQAISAGGENAQAQS 61
5 DNPKTEKAWAIAKTPAIKQHRQQQIETHEHLSALL-EQNGLATSIFFKAGA-----S 58
62 AERV---INQALKKLPSQSPDDIPASSSLIKVIRRAQAQKRGDTHLAVDQIMGLL 118
59 IPRVNDQVNSFTAQPKLSPSESYLGRSLDKLDNAEIAKSKYGGDYISIEHMAAYG 118
119 EDSQI-RDLLENGVATARVSEVKEKRGKESAGDTNFOALTYGRDLVQQA-- 175
119 QDRGLKGLNREIGLTENKLAIEIKQIRGTQ--KVTDQNPGEKYESLEKYGRDLTELARE 176
176 GKLDVIGRDEIRRVRIILSRRTKNNPVLIGPCVGTATVEGQAQRIYKGVDPNSLTD 235
177 GKLDVIGRDEIRVTIQILSRRTKNNPVLIGPCVGTATVEGQAQRIYKGVDPNSLTD 236
236 VRLISLDMGALVAGYRGEFEERLKSVLKEVEDAEAGKVLIFIDEIHLVLAGKTEGSM 295
237 RKLISLDMGALVAGYRGEFEERLKSVLKEVEDAEAGKVLIFIDEIHLVLAGKTEGSM 296
296 AANLKPMALRGALRCIGATILEYRKVKVEKDAFERPQOYVAVPSVPTTISILGLX 355
297 AGNLKPMALRGALRCIGATILEYRKVKVEKDAFERPQOYVAVPSVPTTISILGLX 356
356 EYGEHGVIRQDRALINAQALRYITGRHLPKADLDVDEACANVRVQLDSQPEIDN 415
119 QDRGLKGLNREIGLTENKLAIEIKQIRGTQ--KVTDQNPGEKYESLEKYGRDLTELARE 176

Db 357 ERYVEHGVKIADSAVAALSNRYISDRFLPKALDLDVDEAAAKLMEITSKPELDE 416
Qy 416 LERKRMQLEIELHALEREKDKASKARLIEVRKELDDLRDLKQLPTMKYRKEKRIEIRR 475
Db 417 VDRKIQLENERISLQRENDASAKERLEKLEKELADFKBEQSKLNGQWQSEKTVIQIRT 476
Qy 476 LKQKRELFMSLOAERYDLARAADLRVGAIOEVRSATIAQLGTSSEENV---MLTEN 531
Db 477 VKETIDQVNLIEIQQAQEDVDYKAAELQYKLTDLQROVEALQTLAQOOTSLSLREE 536
Qy 532 VGPEHIAVSVRWGTIPVTRLGQNEKERLIGLADRLHKKRVVQGVQAVNAVSEAILRSRAG 591
Db 537 VLBSDDIAEIIISKVGTIPISKVSEKEKLLHLELHRSRVIGQDEAVTAAVAEQSRAG 596
Qy 592 LGRQAQPTGSLFGPTGVGKTELAKALABOLPDDENLLVRIDMSEYMEKHAVALRMGAP 651
Db 597 LSDPNRPTASFIPLGPTGVGKTELAKALAKNLFDTTEALVRIDMSEYMEKHAVALRMGAP 656
Qy 652 PGYVHEGGQTEAVRRPYPYCVILFDEVEKARHVAFTLLQVLDGRLTDGQRTVDFR 711
Db 657 PGYVHEGGQTEALRRPYPYVILFDEVEKAGHGVNMLQILDDGRLTDGQRTVDFR 716
Qy 712 NSVIMTSLGAEHLIAGLTKVMEVARDCVNRVRRKHFPELNLRLDEIVVDFPLSHD 771
Db 717 NTIIMTSLGQYILDVAGDDSRVEEMRSRVMDVNRNFRPEFLNRVDETIIFHGLQKS 776
Qy 772 QLRKVARLQMKDVAVRLAERGVVALVTDAAALDYILAESYDVPYVGPPIRMMWKKVTEL 831
Db 777 ELRSIVQIQSLATLEBQKLTLDKALDFLAAGVDPYVGPPIRMMWKKVTEL 836
Qy 832 SKOVVREIDENSTVYIDAGADLVYRVESGGIV 865
Db 837 AKGILRGDYKPGETIVVDETDRLSFTSLRGDLV 870

RESULT 5
US-10-047-260-38
Sequence 38, Application US/10047260
Publication No. US20020164706A1
GENERAL INFORMATION:
APPLICANT: Huang, Lisa
APPLICANT: McClusky, Michael
APPLICANT: Larossa, Robert
TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA
FILE REFERENCE: CL1715 US NA
CURRENT APPLICATION NUMBER: US/10/047,260
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/264,925
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Microsoft Office 97
SEQ ID NO 38
LENGTH: 872
TYPE: PRT
ORGANISM: Synechocystis sp. strain PCC6803
US-10-047-260-38

Query Match 50.3%; Score 2288; DB 14; Length 872;
Best Local Similarity 52.6%; Pred. No. 1.7e-154;
Matches 460; Conservative 167; Mismatches 229; Indels 18; Gaps 7;
2 NPEKTHKTNETIATAHVAHAGHQAFTPLHAGALISDPTGIFPQAISAGGENAQAQS 61
5 DNPKTEKAWAIAKTPAIKQHRQQQIETHEHLSALL-EQNGLATSIFFKAGA-----S 58
62 AERV---INQALKKLPSQSPDDIPASSSLIKVIRRAQAQKRGDTHLAVDQIMGLL 118
59 IPRVNDQVNSFTAQPKLSPSESYLGRSLDKLDNAEIAKSKYGGDYISIEHMAAYG 118
119 EDSQI-RDLLENGVATARVSEVKEKRGKESAGDTNFOALTYGRDLVQQA-- 175
119 QDRGLKGLNREIGLTENKLAIEIKQIRGTQ--KVTDQNPGEKYESLEKYGRDLTELARE 176

JLT 6
 00-0-156-761-14770
 Sequence 14770, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
 APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 14770
 LENGTH: 879
 TYPE: PRT

776	WMTS	NLGSEYLVNQ	EGEDTGAV - REQVMGVRAHFRPEFLNRVDEITILFHRLOKSEMR	774
776	VARLQ	KDQVAVRLAERGVALAVTDAALDYII	LAESVDVPYIGARPIRRMEKKVVTLSKMW	835
775	IVDIQ	FARLT	KLLEDKRVLDDAARDWIAEKWDPAYGARPLKRVQORSQDPLAEMI	834
836	VREEIDENSTVIYD	IADAGDLVY	857	
835	LEGSVKDGDHVAIS	AEQGVLT	856	
RESULT 8				
US-10-369-493-19762				

PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20488
LENGTH: 857
TYPE: PRT
ORGANISM: Rhodospseudomonas palustris
10-369-493-20488

Query Match 49.8%; Score 2264.5; DB 12; Length 857;
est Local Similarity 53.1%; Pred. No. 8e-153;
atches 458; Conservative 159; Mismatches 234; Indels 11; Gaps 7;

1 MNPKFTHTKNTIATAHELAVNAGHAQFPLHLGALISDPTGIFFQAISSAGGENAAQ 60
|||:|||:|||:|||:|||:|||:|||:
1 MNVEKYTERVGRGFQSQAQSLAMREGHQQFSPLHILKVLLDDSEGLAGLIDRAGNSRA- 59
|||:|||:|||:|||:|||:|||:|||:
61 SAERVINGCALKKLPSSQ--PPDPDI PASSSLLTKVIRRAAQAKSKSGDTHLAVDQLIMGLL- 118
|||:|||:|||:|||:|||:|||:|||:
60 -ILKATEALCKMFKVSGSGAGQVYLATPAPADPAEKAKEAGDSFVTVERLLLSL 118
|||:|||:|||:|||:|||:|||:|||:
119 -EDSQIRDLLNEVG VATARKVSEVKLGKGKKVESASGDPTNFQALKTYGRDLIVEQA-- 175
|||:|||:|||:|||:|||:|||:|||:
119 DKDSFAGQLLTGGVTPQNIAAINALF--KGR TADSAENA YDALKKYARDLITOAARD 176
|||:|||:|||:|||:|||:|||:|||:

```

176 GKLDVIGRDEIEIRVVILSRRTKNNPVLIGEGVGKTAVVEGLAORIVKGVDPNSLTD 235
177 GKLDVIGRDEIEIRKIITQVLSRRTKNNPVLIGEGVGKTAIVEGLAIRLINGDVPESGLK 236
236 VRLISLDMGALVAGAKYGEFEERLKSVLKEVEDAEKGKVIILFIDEIHLVLGAGKTGESMD 295
237 KKLIALDMGALIAGAKYGEFEERKA VLNEVTAAEGGIILFIDEMHTLVGAGKADGAM 296
296 AANLEKPMLAGOLRCIGATTILEYRKVKVEKDAAFERRPOQVYVAEVSPTTIIILGLK 355
297 ASNLKPLARAGEHICIGATTIDELDKVEKVEDAALARFPQVFVSEPTVEDTVSILGLK 356
356 EKYEGHGVRIQDRALINAAQLSARYITGRHLPKADILDVDEACANVRVOLDSPSEIDN 415
357 DKYEQHGHVRIADSALVAAVLTSLNRYITDRFLPDKAIDLMDBAARLKMQVDSPKEULDS 416
416 LERKRMQLEIHUALERKDKASKARLIEVRKELDURDKLOPTIMKYRKEKERIDIRR 475
417 MDREIVRLKIEQEALKKETDPGSKARLVTLEKELADLEKSAALTQRWSAEKNLSDAQ 476
476 LKQKREELMFSLQEAEREYDLARADLEYGAIQEVESAIAOLEGTSSSEENVMLTENYGPE 535
477 LKSELDAURLIELANAQRGEYQRAGETAYGRIPLEKKIAEIE-ANNSGAMWEEAVTAN 535
536 HIAEVVSWTGIPVTRLQGNKEKRLIGLADLRHKRVVGQONAVNAVSEAILRSAGLGRA 595
536 HIAQVVSMTGPVDDMKLEGEKEKLLRMEEQLGQRVVGQFEAVHAVSTAVRRARAGLQDP 595
596 QOQTGSFLFGPTGVGKTELAkalaEQLFDDEENLLVRLDMSEYWEQHSVRELICAPGYV 655
596 NRPMSGFMLGPTGVGKTELTKALAELYFDDETAMVRLDMSEFMEKHISVARLIGAPPYV 655
656 GHEGGQLTEAVRRPYCVILFDEVEKAHVAFNTLLQVLDGRLTDQGRTVDFRNSVI 715
656 GYDEGGVLTEAVRRPYQVILFDEIEKAHPDVFNVLQVLDGRLTDQGRTVDFRNLI 715
716 IMTSLNGABHLAGITGKVTMVEARDQVMREVRKHFRPELLNRLEIDVFPDLSHDQLRX 775

```

477 VKKEIRVNLEIQOARDYDLNRAAEALKYGNLTSILHQLAEVAEALASQKSGKSLYGE 536
532 VGPEHIAEVVSRWTGIPVTELGONEKERLIGLADRLHKKRVGQONAVNAVSEALERSAG 591
537 VTEADIAEILSKWTGIPISKLVESEKEKLHLEDELRHVRVIGOREAVTAVADAIQSRAG 596
592 LGRAOQPTGSLFELGPTGVGKTELAKALAEQDDENLVLRIDMSEYMEQHSVSRILIGAP 651
597 LADPNRPIASFELGPTGVGKTELAKALAGYMFDSALVRIDMSEYMEKHAVALRILIGAP 656
652 PGVGHGEGQLTEAVRRPYCVILFDEVEKAHVAVNTLLQVLDDGRLTDGGRVTVDFR 711
657 PGVGVGEGQLTEARRRPYSVILFDEIEKAHPDVENIFLQILDDGRVTVDAQGHKVDK 716
712 NSVIIMTSNLGAHLLAGLTGKVTMEVARDVCMREKHFPRPELLNRLDEIVVDFPLSHD 771
717 NALLIMTSNLGSIYLDVAGDNHAYDEMRRRVMEANRSPRPFELNRIIDBIIFHGLDKK 776
772 QLRKVARLQMKDVAVLAERGVAVTDAALDYILAESYDPVYGPARRRMEKKVVTTEL 831
777 ELRQVILQVORLQRORLADRKSILKLSDAALDFLAEGVDPVYGARPLKRAIQRELETOI 836
832 SKMVRREIDENSTVID 849
837 AKAILRGEFTDGNITFVD 854

LT 9
0-369-493-20953
Sequence 20953, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20953
LENGTH: 862
TYPE: PRT
ORGANISM: SYNECHOCOCUS SP. WH 8102
0-369-493-20953

Query Match 49.3%; Score 2243; DB 12; Length 862;
Best Local Similarity 52.7%; Pred. No. 2.7e-151;
Matches 455; Conservative 166; Mismatches 221; Indels 22; Gaps 10;

1 MNP-EKFTKNTIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGENA 58
1 MQTARQTEQAAVAIAVAQAASHQOLETEHLLALLRQ-NGLAGRLSKTGV-- 57
59 AQSAERVINGALKLPSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL 118
58 VTTFEASVEGHLQRLPSGAPDSVFLGRSLKALDRAEQORDGFGDSFIAIEHLALLA 117
119 EDSQI-RLLANEVGVATARKVSEKLRGKGGKVESASGDTNFQALKTVGRDLVEQA-- 175
118 EDDRCGRQLSQAGVTNTLKEAITAVRGNQ--TVTDQNPEATYSLAKYGRDLTAARD 175
176 GKLDPVIGRDEEIRRVIRLSRTKNPNVLIGPVGKTAIVGELAQKRVKGVNLSLTD 235
176 GQLDPVIGRDEEIRRVIRLSRTKNPNVLIGPVGKTAIVGELAQKRVKGVNLSLTD 235
236 VRLISLDMGALVAGAKYRGFEERLKSVLKEVEDAEGKVFILFIDEIHLVLAGKTEGMD 295

236 RQLITLDMGALVAGAKYRGFEERLKAIVLKEVTTSDGQIVLFIDEIHTVVGAGSGAMD 295
236 AANLFKPMIARGLRCIGATTLEBYRKVYKEDAAFERFOQVYVAERSVPTTISILRLK 355
236 ASNLKPMIARGLRCIGATTLEBYRKVYKEDAAFERFOQVYVAERSVPTTISILRLK 355
336 EKYEGHGVRIQDRALINAAQISARYITGRHLDPKADILVDEACANVRVQLDSDPEEIDN 415
336 ERYEVHGVRIADSNALVAAMLSRYITDRFLPKADILVDESAARLKEWITSPEQIDE 415
416 LERKMQLEIELHALEREKDKASKARLIEVKEKLDLDFKLQPLTMKYRKEKERIDEIR 475
416 IDRKTILQEMEKLSIGRESDSASOERLQIERELAEELGEQSSNAQWQSEKGAIDQLSA 475
476 LKQKEBELMPSLQEAERHYDLARAADILRYGA-----IOVESALIAQLECTSSSEENVML 528
476 LKEETERVQLQVEQAKRNYDLNKAEEYGLTALQRLQEQEDELLEDEDTG---DKTLL 532
529 TENVPREHIAEVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVGQONAVNAVSEALERS 588
533 REEVTEDDIAEVIKWTGIPVARLVQSEMEKLLQLEDDLHQRVIGQOAVTAVADAQORS 592
589 RAGLGRAOQPTGSLFELGPTGVGKTELAKALAEQDDENLVLRIDMSEYMEQHSVSRIL 648
593 RAGLSDPNRPIASFELGPTGVGKTELAKALAEQDDENLVLRIDMSEYMEKHTVSRIL 652
649 GAPGVGHGEGQLTEAVRRPYCVILFDEVEKAHVAVNTLLQVLDDGRLTDGGRTV 708
653 GAPGVGHGEGQLTEAVRRPYCVILFDEVEKAHVAVNTLLQVLDDGRLTDGGRTV 712
709 DFRNSVIIMTSNLGAHLLAGLTGKVTMEVARDVCMREKHFPRPELLNRLDEIVVDFP 767
713 DFTNVLTLTSNIGSQSILE-LAGDPQOHTAMEQVNEALKAKPERPELNLDDQIIFRS 771
768 LSHQRLKVARLQMKDVAVLAERGVAVTDAALDYILAESYDPVYGPARRRMEKKV 827
772 LEKSELRRIVSLQVERLSRLEORKLQILSDIAADMLATIGFDPVYGARPLKRAIQREL 831
828 VTELKMKVREIDENSTVIDAG 851
832 ETPIAKAILAGLSEGGTVQVDAG 855

RESULT 10
US-10-369-493-22555
Sequence 22555, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 22555
LENGTH: 905
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22555

Query Match 49.1%; Score 2234.5; DB 12; Length 905;
Best Local Similarity 49.5%; Pred. No. 1.2e-150;
Matches 460; Conservative 158; Mismatches 258; Indels 53; Gaps 11;

6 FTHKNTIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGENA 61

6 FTDKAAKTLSDAYIAQSYGHSQITPHIAAALLSDSDSNGTTLRLTRIVDKAGD--GQK 63
62 AERVINOALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLEDS 121
64 FERSVTSLSVLPQADPPPEQVTLSPESAKLLRNAHELQTKQDSYIAQDHFIAVFTKDD 123
122 QIRDLLENGVATARVSEVEKLKRGKGVKVESAGDTNFQALTKYGRDLVEQA--CKLD 179
124 TKSLAAGVTPKAFEFANVRG--NKRIDSQNAEEGFDALNKFTVDTLTELARNQOLD 181
180 PVIGREIRVRVRLSRTKNNPVLIGEPGVGTAVVGLAQRIVKGDVPSNLTVDRLI 239
182 PVIGREIRVRVRLSRTKNNPVLIGEPGVGTAVVGLAQRIVKGDVPSNLTVDRLI 241
240 SLDMGALVAGAKYGEFEERLKSILKEVEDAEGKVILFIDIEHLVLAGK--TEGSDMAAN 298
242 SLDVGLVAGSKFRGEFEERIKSVLKEVESETPILFVDEMHLLMGAGSGEGGMDAAN 301
299 LFKPMARGQLRCIGATTLBEYRKVVEKDAAFERRFOQVYVAEVSVDITISILGLKEKY 358
302 LKPMARGQLKHCIGATTLBYKVKYKDAFAFRFQIILVKEPSIEDTISILGLKEKY 361
359 EGHGVRIQDRALINAAQARSYITGRHLPDKAIDLVDDEACANVRVOLDISOPEIDNLER 418
362 EVHGVVTSIDRALVTAHLASRYLTSRLPDSADLDVDEAAAARVTVRESQPEVLDNLER 421
419 KRMQLEIEHLALEREKQKASKARLIEVKEIDLDLRLKLOPTMTKYRKEKERIDEIRLKQ 478
422 KLRQVRVIRALEREKDEASKERLKAARKEAEQVEETPRIRREYKELEKSRGSELOAKR 481
479 KREELMFSLOEAREVYDLARADIRYGAIOVESAIQLEGTSSEENV-----526
482 RLDELKAKAEDARRNDTFLAADLKYYGIPDQKRIEYLEQKQKADABAFANAQPSSEP 541
527 MLTENVGPEHIAEVVSRWTGIPVTRLGQNEKERLIGLADRLHKRVGQONAVNAVSEAIL 586
542 LLDVVGPDQINEIIVARWTGIPVTRLTKTERULNMEKVLKQVIGQNEAVTAVANAI 601
587 RSRAGLRAQOQTGSLFGLPTGVGKTELAKALABQLFDDENLLVRIDMSYMEQHSVSR 646
602 LSRAGLSDPNQPIASFLCPGSPGSGTKLTLLKALASFMFDDENAMIRIDMSYMEKHSVSR 661
647 LIGAPGVGHEEGQLTAEVRRPVCVILFDEVEKAVHVAVENTLLQVLDGRLTDGGR 706
662 LIGAPGVGHEAGQLTAEVRRPVCVILFDEVEKAVHVAVENTLLQVLDGRLTDGGR 721
707 TVDERSVIMTSNLGAELHLLA--GLTKGVMEVARDVREVRKHPRELINRLDEIV 763
722 VVDKNAVIMTSNLGAELVITDNESSDDGKID--STTREMVMNSIRGFRPEFLNRISIV 780
764 VPDPLSHDLQKRVARLQMDVAVRLA--ERGVALAVTDAALDYLLAESYDVPYGARPIR 821
781 IFENRRLVRDIRNIVENRILEVQKRNHRSIKIEVDEAKDLGSGAGSYPAYGARPIR 840
822 WNEKQVTELSKVMREEDENSTVYIDAGADLVVRVESGGLVDASTGCKSKDVLILHAN 881
841 VTQONVLPMAVLIINGQLRDKETAHVVVQNGKI FVKN-----HEAN 883
882 GPKRSDAAQAVKVMRIEIEDDDNEEMIE 910
884 ANGSAD-----IDMDGIDDVNDDEELE 905

SULT 11

Sequence 11982, Application US/10369493
Publication No. US2003023675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10369,493
;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 11982
;; LENGTH: 868
;; TYPE: PRT
;; ORGANISM: Mesorhizobium loti
US-10-369-493-11982

Query Match 48.9%; Score 2225; DB 12; Length 868;
Best Local Similarity 52.0%; Pred. No. 5.3e-150;
Matches 448; Conservative 159; Mismatches 245; Indels 10; Gaps 6;
QY 1 MNPEKTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPOAISSAGGENAAQ 60
DB 1 MLEKYSERVFGIQAQTMALSRNHQFTPEHLKVLVDDEGLAASLIERAGN--VR 58
QY 61 SAERVINOALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL-- 118
DB 59 DVKLGVTALEAMPKVBEGNGQLYLAQPLAKVFAEELAKKAGDSFVTVERLLQALTWE 118
QY 119 ESQIRDLLENGVATARVSEVEKLKRGKGVKVESAGDTNFQALTKYGRDLV--EOAG 176
DB 119 KSAKTADIIKAGVTAQALNQVINDVR--KGTADASAEQNTDALKKYARDLTADARAG 176
QY 177 KLDPVIGRDEIRRVVRLSRTKNNPVLIGEPGVGTAVVGLAQRIVKGDVPSNLTVD 236
DB 177 KLDPVIGRDEIRRVVRLSRTKNNPVLIGEPGVGTAVVGLAQRIVKGDVPSNLTVD 236
QY 237 RLISLDMGALVAGAKYGRGEFEERLKSILKEVEDAEGKVILFIDIEHLVLAGKTSMDA 296
DB 237 QLMALDVGALLIAGAKYGRGEFEERLKSILKEVEDAEGKVILFIDIEHLVLAGKADGMDA 296
QY 297 ANLKPMLARGQLRCIGATTLBEYRKVVEKDAAFERRFOQVYVAEVSVDITISILGLKE 356
DB 297 SNLLKPMALARGELHCVGATTLBEYRKVVEKDAAFERRFOQVYVAEVSVDITISILGLKE 356
QY 357 KYEGHGVRIQDRALINAAQARSYITGRHLPDKAIDLVDDEACANVRVOLDISOPEIDNL 416
DB 357 KYEGHGVRIQDRALINAAQARSYITGRHLPDKAIDLVDDEACANVRVOLDISOPEIDNL 416
QY 417 ERKMQLEIEHLALEREKQKASKARLIEVKEIDLDLRLKLOPTMTKYRKEKERIDEIRL 476
DB 417 DRRIMQKIEREALKVETDDASKORLVRLKELVGLLEESTETITAKWQAEKOKLGLAADL 476
QY 477 KOKREELMFSLOEAREVYDLARADIRYGAIOVESAIQLEGTSSEENVMLTENVGPEH 536
DB 477 KOKREELMFSLOEAREVYDLARADIRYGAIOVESAIQLEGTSSEENVMLTENVGPEH 536
QY 537 IAEVSRWTGIPVTRLGQNEKERLIGLADRLHKRVGQONAVNAVSEAILRSRAGLGRAQ 596
DB 536 VAMIVSRWTGIPVDMKQGERDKLRLMEDEIGKRVVVGQGEAVQAVKAVRRARAGLDQFN 595
QY 597 OPTGSLFGLPTGVGKTELAKALABQLFDDENLLVRIDMSYMEQHSVSRILGAPGVYG 656
DB 596 RPTGSLFGLPTGVGKTELAKALABQLFDDENLLVRIDMSYMEQHSVSRILGAPGVYG 655
QY 657 HEEGGQLTAEVRRPVCVILFDEVEKAVHVAVENTLLQVLDGRLTDGGRITVDYFRNSVII 716
DB 656 YEGGALTEAVRRPVCVILFDEVEKAVHVAVENTLLQVLDGRLTDGGRITVDYFRNSVII 715
QY 717 MTSNLGAELHLLA--GLTKGVMEVARDVREVRKHPRELINRLDEIVVFPDPLSHDLQKRV 776
DB 716 MTSNLGAELHLLA--GLTKGVMEVARDVREVRKHPRELINRLDEIVVFPDPLSHDLQKRV 774
QY 777 ARLOMKDVAVRLAERGVALAVTDAALDYLLAESYDVPYGARPIRREVRKVMKVVTELSKVV 836
DB 775 VEIQFKRLSLLVDRKITUSLDHEALTEWLAAGKYDPAYGARPIRVMQKELQDPLAEKIL 834

; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14088
 ; LENGTH: 861
 ; TYPE: PRT
 ; ORGANISM: Salmonella typhi
 ; NAME/KEY: VARIANT
 ; FEATURE:
 ; LOCATION: (1)...(861)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-09-815-242-14088

Query Match 48.4%; Score 2204; DB 9; Length 861;
 Best Local Similarity 52.1%; Pred. No. 1.7e-148;
 Matches 449; Conservative 155; Mismatches 244; Indels 14; Gaps 9;

QY 1 MNPEKFTKTTTATAHELAVNAGHAQFTPLHLAGALISDPTGTFPQAOISSAGGENAAQ 60
 DB 5 MRLDLDTKNFQALADAQSLAGHDNQFIEPLHMSALINOGGSIPLITSA-GINAGQ 63
 QY 61 SAERVINOALKKLSQSPDPDDI1PASSSLIKVIRRAQAQKSRGDTHLAVDQLMGLLED 120
 DB 64 -LRTAIDQALSSELPOVEGTGGDVQPSSELVRLNCDKLAQKRGDNFISSELVLALES 122
 QY 121 -SQIRDLENGVATARVYKSEVKLRGKKGKVSASGDTNFQALKYTGRLDLEQA--GK 177
 DB 123 RGTLDLLKSGAGTANTITQATIEOMRG--GESVNDQGAEDQORQALKYTVDLTERAOGK 180
 QY 178 LDPVLGRDEIRRVRIILSRRTKNNPVLIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 237
 DB 181 LDPVLGRDEIRRTIQVLRRTKNNPVLIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 240
 QY 238 LLSLDMGALVAGAKTRGFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTEGSDAA 297
 DB 241 VLALDMGALVAGAKTRGFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTEGSDAA 300
 QY 298 NLPKPMALGQURCTGATTLBEYRYKYVEKDAAPERRFOQVYVVEPSVPTDIIIRGLKEK 357
 DB 301 NMLKPALARGELHCVGATTLDEYROYIEKDAALERRFOQVYVVEPSVPTDIIIRGLKEK 360
 QY 358 YEGHGVRIODPALNAAQLSARYITGRHLDPKADLVDEACANRVOLDSPQPEIDMLE 417
 DB 361 YELHHVQITDPAIVAAATLSHRYIADRLPKADLVDEACANRVOLDSPQPEIDMLE 420
 QY 418 RKRMOLEIEELHALEREKOKASKARLIEVRKELDDLDKLOPLTKYRKEKERIDEIRLKL 477
 DB 421 RRIQIKLEQALMKESDEASKRLDMLNEELDDKEROYSELEBWEKAKASLSGTQTIK 480
 QY 478 QKREELMFSLQBAERYDLARAADLRPAIYQVESAIAQLE-GTSSSENVN--LTENVGP 534
 DB 481 ASLEQAKIAIEQARRVGLDARMSELQYKIPLEK---QLEAATQSEKTRMLRNKRVTD 537
 QY 535 EHIAEVVSRWGTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVEAILRSRAGLGR 594
 DB 538 AEIAEVRLRWGTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVEAILRSRAGLGR 597
 QY 595 AQOPTGSFLPGTGVGKTELAKALAEOLFDDENLLVRIDMSSEYMEQHSVSRKICAPPGY 654
 DB 598 PNRPIGSFLPGTGVGKTELCKALANFMFDDSDAMVRLDMSEFMEKHSVSRKICAPPGY 657
 QY 655 VGHEGGQQLTEAVRRRPYCVIILFDEVEKAHVAFTNLLQVLDGRLTDGQRTVDFRNSV 714
 DB 658 VGVEGGYLTTEAVRRRPYCVIILFDEVEKAHVAFTNLLQVLDGRLTDGQRTVDFRNTV 717
 QY 715 IIMTSLNGAEHLILAGLTGKVTMEVARDVCMREVRKHPPELLANRLDEIWFDPISDOLR 774
 DB 718 VIMTSLNGSD-LIQERFGLDYGRMKEMVGVVVSQNFPEFNRIDFVVVHFPLGEQHIA 776
 QY 775 KVARLOWKDVAVLAEARGVALAVTDAALDYILAESYVDPVYGARPIRRRMWEKKVVTLSKM 834
 DB 777 STIAQIQRLRYKELEERGVEIHSDEALKLISANGYDVPYGARPLKRAIQIQIENPLAQQ 836

345 PDTIILRLKEKVEGHGVRIODRALINAAQLSARYITGRHLDPKADLVDEACANRV 404
 357 SETIILRLKEKVEGHGVTSIDAAIVAAANLAARVLTSSRLPDSADLIDIDEEAAVAV 416
 405 QLDSPQPEIDNLERKRWQLRIEELHALEREKOKASKARLIEVRKELDDLDKLOPLTKYR 464
 417 ARESPQPEIDNLERKRWQLRIEELHALEREKOKASKARLIEVRKELDDLDKLOPLTKYR 476
 465 KEKERIDEIRLKKRRELMSLOEARRRYDLAARADLRPAIYQVESAIAQLE----- 518
 477 QEQORAKALQBARWKLSELRQKABEASRMGDHRSRAADLQYVAIPEQAVIKRLEKEKAAA 536
 519 -----GTSSENVNLTENVGPPIHAEVVSRTGIPVTRLGQNEKERLIGLADRLHRRV 572
 537 DAALNAAAEATGGAMITDVVGPDOINEIVARWTGIPVTRLTSEKELHMEKHLKIVV 596
 573 QONQAVNAVSEAILRSRAGLQRAOQPTGSFLPGTGVGKTELAKALAEOLFDDENLLVR 632
 597 GQKEAVQSVNAILRQSLGNPNQPP-SFLFCPSGTGKTLITKALAEFLFDDPKAMIR 655
 633 IDMSSEYMEQHSVSRKICAPPGYVGHGGQQLTEAVRRRPYCVIILFDEVEKAHVAFTNLL 692
 656 FMSBYQERHLSRMIGAPGYVGHGGQQLTEAVRRRPYCVIILFDEVEKAHVAFTNLL 715
 693 QVLDGRLTDGQRTVDPRNSVIMTSLNGAEHLILAGLTGKVTMEVARDVCMREVRKH 750
 716 QLMDDGRLTDGQRTVDPRNSVIMTSLNGAEHLILAGLTGKVTMEVARDVCMREVRKH 774
 751 FRPELNLRLDEIVDPDPLSHDOLRKVARLOWKDVAVRIA--ERGVAVAVTDAALDYILAE 808
 775 FLPEFLNRISIVIFENLRTRIRKIVELRAETIQRLQDNRNVKIEVSEAKDKLIGAL 834
 809 SYDPYVYGARPIRRRMWEKKVVTLSKMVREIENSTVYIDAGADLVYRVESGLVDAS 868
 835 GYSPAYGARPLQVRLEKSVLRLAVLILRGSRD-----GEVARVVQGGKITV- 893
 869 TGKSDVLIHLIANGPKRSDAAQAVKMKRIETIEDDDNEEMIED 911
 884 -----LPNHP-----EVNDEDEMWLDE 901

JUL 14
 09-815-242-14088
 Sequence 14088, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308

835 VVREEIDENSTVYIDAGADLV 856
:: : : : :
837 ILSGELVPGKVRLEANDDRIV 858

JUL 15 10-369-493-10608
sequence 10608, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10608
LENGTH: 858
TYPE: PRT
ORGANISM: Sphingomonas aromaticivorans
10-369-493-10608

Query Match 48.3%; Score 2199.5; DB 12; Length 858;
Best Local Similarity 52.4%; Pred. No. 3.5e-148;
Matches 454; Conservative 160; Mismatches 232; Indels 21; Gaps 9;

1 MNPEKFTHTNETTATAEHLAVNAGHAGFTPLHLAGALISDPTGIPPOAISSAGGENAAQ 60
1 MNLEKFTDRAGKFLQAQTVAIRNHHORITPDHILKALLEDESEGMAISGLIQRAGG-NAA- 58
61 SAERVINOALKKLP-----SQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLI 114
59 LAQTEVDKALAKIPAVSGSQAQATPGLD---NDARVLDLSAEQIAAKSNDSFVTVERML 114
115 MG--LLEDSQIRDLLNEVGATARKVEKLRGKQKVESASGDTNFQALKTYGHDLV 172
115 VALTIATTSAGQALKAAANTVAQALEAATQLRG--GRTADSASAEANAYDAMKKYARDLT 172
173 EQA--GKLDPVIGRDEERRVRIILSRRTKPNVLIPEPGVGKTAUVVEGLAQRIKVGDDVP 230
173 EAAREGKLDPVIGRDEIRRVQILARRTKNNPALICEPGVGKTAIAEGLALRIANGDVP 232
231 NSLTDVRLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVILFIDHILVLGAGKT 290
233 DSLKDRRLMALDMSLIAGAKYRGEFEERLKAVIDEVKGAEGEILFIDEMHTLIGAGKS 292
291 EGSMDAANLKPMLARGOLRCIGATTLEEYRKYVEKDAAPERFQOYVVAEPSVPTISI 350
293 EGAMDAGNLLKPALARGELHCIGATTLEDEYQYKDPALQRRFPVPVGEPTVEDTISI 352
351 LRGLKEKYEGHGVRIODRALINAQLSARYITGKHLPKAIDLVDACANVRVOLDSP 410
353 LRGIKDKVELHHGVRIADNAIVAATLSNRVYISDRFLPKAIDLMDAASRIRMEVESKP 412
411 BEINDLERKQMLEIEHLERFKDKASKARLIEVRKELDDLRKQLQPLTKYRKEKERI 470
413 EEIEKLDRIIQQMKIEEMALAKETDQASKDLATLREELANQEQOQSAELTTRWQNERDKI 472
471 DEIRLQKRELMEFSLQAEERYDLARAADLRYLGAIOEVESAIAQLEGTSSEENVMLTE 530
473 AAEQVKKALDAARSELEVAQRNGDLAKAGELAYGRIPELERQLAEAGVGS--QNAMLRE 530
531 NVGPEHIAEVVSRWTGIPVTRIGQNEKERLIGLADRLHKRVVQGNQAVNVAISAILRSRA 590
531 EVTAEDIAAVVSKWTGVPDRMMEGEREKLLHMEALGKRVIGQKDAVLAVSKAVRARA 590

QY 591 GLGRAQQPTGSEFLFGPTGVGKTELAKALAEQLFDDENLLVRIDMSYMEQHSVSRLLIGA 650
Db : : : : :
591 GLQDPNRPLGSFLFGPTGVGKTELTKALAGFLFDDDNAMVRIDMSSEFMEKHSVSRLLIGA 650
QY 651 PFGYVGHBEQGOITAEVRRRPPYCVILFDEVEKAVAVFNTLLQVLLDDGRLTDGQRTVDF 710
Db : : : : :
651 PFGYVGYDEGGVLTETAIRRRPYQVVLFDEVEKASDVFNVLQVLLDDGRLTDGQRTVDF 710
QY 711 RNSVITMSNLGAELHLAGLTGKVTMEVARDCVMEVRKHPPELNLRLDEIVVDFPLSH 770
Db : : : : :
711 TWTLILITSLNSQY-LANLEGGQVQSVPEQVMDVRAHFRPEFLNLRLDEILFHLGH 769
QY 771 DQLRKVARLQMKDVAVRLAERGVALAVTDAALDYILABSYPVYGARPIRRMMEKKVTE 830
Db : : : : :
770 EEMAPIVDIQVGRVANLLKDKRIVLDTDAAKRWLGRVGYDPVYGARPLKRAVQRYLQDP 829
QY 831 LSKWVVRBEIDENSTVYIDAGADLVY 857
Db : : : : :
830 LAEKLIGGEVDPDGTVRIDEQDGLSF 856

Search completed: February 13, 2004, 01:43:03
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:25:21 ; Search time 770 seconds
(without alignments)
10885.396 Million cell updates/sec

le: US-09-812-350-30

fect score: 3105

quence: 1 aaagtatacaattttacac.....tctttgggttttttcata 3105

oring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

arched: 2552756 seqs, 1349719017 residues

al number of hits satisfying chosen parameters: 5105512

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase :

N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3105	100.0	3105	22	Arabidopsis thalia
2	2732.8	88.0	2736	24	Arabidopsis thalia
3	2732.8	88.0	2736	25	Arabidopsis thalia
4	2535.8	81.7	6375	22	Arabidopsis thalia
5	1638.4	52.8	3049	22	Glycine max heat s
6	1564.4	50.4	3052	22	Nicotiana tobacum
7	1345	43.3	3084	22	Zea mays heat choc
8	1334.2	43.0	3058	22	Triticum aestivum

Triticum aestivum
Zea mays 101 kDa h
Zea mays heat choc
Phaseolus lunatus
Triticum aestivum
Enterococcus faec
Enterococcus faec
Listeria innocua c
Listeria innocua D
Synchococcus CLPB
Synchococcus sp h
Genomic sequence o
Listeria monocytog
Lactococcus lactis
Haemophilus influ
Haemophilus influ
Staphylococcus aur
Staphylococcus aur
Trypanosoma brucei
High level promote
Leishmania major h
Staphylococcus aur
E. coli DNA for ce
E. coli clpB gene.
Leishmania donovan
Genomic fragment #
Salmonella typhi D
Staphylococcus epi
C glutamicum codin
Corynebacterium gl
C glutamicum codin
S. epidermidis gen
C. trachomatis gen
Complete genome se
C. pneumoniae DNA
Nucleotide sequenc
M. capsulatus gene

ALIGNMENTS

RESULT 1

AAI66065

ID AAI66065 standard; DNA; 3105 BP.

XX AAI66065;

AC AAI66065;

XX 14-JAN-2002 (first entry)

DT Arabidopsis thaliana heat shock protein 101 gene (GenBank: U13949).

DE Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

KW Arabidopsis thaliana; ds.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN WO200170929-A2.

XX 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US08836.

XX 20-MAR-2000; 2000US-190769P.

PR 18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

PI Lindquist S, Queitsch C, Vierling B;

XX WPI; 2001-639123/73.

DR P-PSDB; AAM51659.

661 TATGGAAGAGATTGGTTGAGCAAGCAGGGAAGCTTGATCCTGTGATTTGGTCGTGATGAG 720

1741 ATGCTCAGAGAAACGTTGGCCCTGAACACATTCCTGAGGTTGTGAGCCGTTGACAGGG 1800
 1801 ATTCAGTACGAGACCTTGGCCAAATAGAGAGAGAGTTGATTTGCTTGTCTGATAGG 1860
 1801 ATTCAGTACGAGACCTTGGCCAAATAGAGAGAGAGTTGATTTGCTTGTCTGATAGG 1860
 1861 TTGCATAGCGGTTGTGGGACAGAACTCAAGCGTAAATGCGATTTCTGAGGCAATTTCTA 1920
 1861 TTGCATAGCGGTTGTGGGACAGAACTCAAGCGTAAATGCGATTTCTGAGGCAATTTCTA 1920
 1921 AGCTCAAGCGGACGACTTGTGAGGSCAACAGCCCACTGATCATTTCTTATTCCTTGGG 1980
 1921 AGCTCAAGCGGACGACTTGTGAGGSCAACAGCCCACTGATCATTTCTTATTCCTTGGG 1980
 1981 CCAACTGGTGTGGCAAACTGAGCTGCCAAGGCTCTTGTGAGCAGCTTTTGTATGAT 2040
 1981 CCAACTGGTGTGGCAAACTGAGCTGCCAAGGCTCTTGTGAGCAGCTTTTGTATGAT 2040
 2041 GAAACCTCTTAGTTGGATGATGATGCGGAATATATGGAACAACACTCTGCTCTCGC 2100
 2041 GAAACCTCTTAGTTGGATGATGATGCGGAATATATGGAACAACACTCTGCTCTCGC 2100
 2101 CTCAATGGGCAACACAGGCTATGTTGTCACGAGGAGGTGACAACTAACTGAGGCT 2160
 2101 CTCAATGGGCAACACAGGCTATGTTGTCACGAGGAGGTGACAACTAACTGAGGCT 2160
 2161 GTGAGGAGGACCTTATTTGTCTATCTCTTTGATGAAGTGGAGAGGCTCATTTGCT 2220
 2161 GTGAGGAGGACCTTATTTGTCTATCTCTTTGATGAAGTGGAGAGGCTCATTTGCT 2220
 2221 GTCTTCAACCTCTGCTCCAGATTTGATGATGATGATGATGATGATGATGATGATGAT 2280
 2221 GTCTTCAACCTCTGCTCCAGATTTGATGATGATGATGATGATGATGATGATGATGAT 2280
 2281 ACAGTCGATTTTCAAGCTCGCTGATATATGATGATGATGATGATGATGATGATGATGAT 2340
 2281 ACAGTCGATTTTCAAGCTCGCTGATATATGATGATGATGATGATGATGATGATGATGAT 2340
 2341 CTTCAGGCTTAACTGGGAAAGTAACTGAAGTGGCCCGGAGCTGTGTGATGCGGAG 2400
 2341 CTTCAGGCTTAACTGGGAAAGTAACTGAAGTGGCCCGGAGCTGTGTGATGCGGAG 2400
 2401 GTGAGGAACACTTTCAGACAGAGCTCTTGAACAGGCTTGAACAGGCTTGTGTGATGCG 2460
 2401 GTGAGGAACACTTTCAGACAGAGCTCTTGAACAGGCTTGAACAGGCTTGTGTGATGCG 2460
 2461 CCCCTTTCACATGACCTTTCAGGAAAGTACCTCGCTTCAATGAAAGAGCTTGTGTGTC 2520
 2461 CCCCTTTCACATGACCTTTCAGGAAAGTACCTCGCTTCAATGAAAGAGCTTGTGTGTC 2520
 2521 CGGCTTGTGAAAGAGGAGTTGCTTGGCAGTCACTGATGCTGCTTGGCAGTATATCTTG 2580
 2521 CGGCTTGTGAAAGAGGAGTTGCTTGGCAGTCACTGATGCTGCTTGGCAGTATATCTTG 2580
 2581 GCAGAGCTTATGACCGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
 2581 GCAGAGCTTATGACCGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
 2641 GTGCTGACAGAACTGCTCAAGATGCTTGTGCTGAGGAAATCGATGAAATCTCCACTGTT 2700
 2641 GTGCTGACAGAACTGCTCAAGATGCTTGTGCTGAGGAAATCGATGAAATCTCCACTGTT 2700
 2701 TACATAGATGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
 2701 TACATAGATGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
 2761 GCTTCAACAGGCAAGATCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
 2761 GCTTCAACAGGCAAGATCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
 2821 GATGAGCTTCAAGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATG 2880
 2821 GATGAGCTTCAAGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATG 2880

QY 2881 GAAATGATCGAGATTAAGTCTTGAACCTCCAAAGTGTGAGTCTTTGAGGTGTGGA 2940
 Db 2881 GAAATGATCGAGATTAAGTCTTGAACCTCCAAAGTGTGAGTCTTTGAGGTGTGGA 2940
 QY 2941 TTGTATGACTTTCAGCTTTCATGCTCTGTTTGGTGTGACTCTTTTGGTGTGACTATAATG 3000
 Db 2941 TTGTATGACTTTCAGCTTTCATGCTCTGTTTGGTGTGACTCTTTTGGTGTGACTATAATG 3000
 QY 3001 GTGTGAACCTGTAAGAACTTACCTTATTTATGTTTTCGTATATGTAATCAATTTTACAT 3060
 Db 3001 GTGTGAACCTGTAAGAACTTACCTTATTTATGTTTTCGTATATGTAATCAATTTTACAT 3060
 QY 3061 AGAAATCCATTGTAGCTCCCAATAATCTTTTGGGTGTTTTCATA 3105
 Db 3061 AGAAATCCATTGTAGCTCCCAATAATCTTTTGGGTGTTTTCATA 3105

RESULT 2
 ABZ12430
 ID ABZ12430 standard; DNA; 2736 BP.
 XX
 AC ABZ12430;
 XX
 DT 21-JAN-2003 (first entry)
 XX
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 235.
 XX
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200216655-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26685.
 XX
 PR 24-AUG-2000; 2000US-227866P.
 PR 26-JAN-2001; 2001US-264647P.
 PR 22-JUN-2001; 2001US-300111P.
 XX
 PA (SCRI) SCRIPPS RES INST.
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 PI Harper JF, Kreps J, Wang X, Zhu T;
 WPI; 2002-304127/34.

Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses
 Claim 144; SEQ ID NO 235; 577pp + Sequence Listing; English.
 The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:
 (a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and
 (b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used in methods of the invention.
 Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Query Match 88.0%; Score 2732.8; DB 24; Length 2736;
 SQ Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;

1021	GAGCGAAGTGTCGTGACACCAATTAGTATCTCTTAGAGACCTCAAGAGGAAAGTATGAGGGA	1081
1243	CATCATGTGTGCGAATCCAGACAGAGCTCTTATAATAGTGTGCTCAGCTGTCTGTCTCGT	1302
1081	CATCATGTGTGCGAATCCAGACAGAGCTCTTATAATAGTGTGCTCAGCTGTCTGTCTCGT	1140
1303	TACATAAATCGTCTGGCAATTTACCGGATAAAGCAAATTGAATTTGGTTGATGAGGCTTGTGGC	1362
1141	TACATAAATCGTCTGGCAATTTACCGGATAAAGCAAATTGAATTTGGTTGATGAGGCTTGTGGC	1200
1363	AATGTGAGATCCAGCTTGATAGTCAACTGAAGAGATTGATACTTTGAAAGGAAGAGG	1422
1201	AATGTGAGATCCAGCTTGATAGTCAACTGAAGAGATTGATACTTTGAAAGGAAGAGG	1260
1423	ATGCAGCTGGAAAATTGAACCTTCACGCCCTTGGAAAAGGGAAGGATAAAGCCAGCAAAAGCT	1482
1261	ATGCAGCTGGAAAATTGAACCTTCACGCCCTTGGAAAAGGGAAGGATAAAGCCAGCAAAAGCT	1320
1483	CGACTTATAGAGGTGCGGAAGAGCTTGATGACTGTGAGGACAGACTTCAGCTCTCTCACG	1542
1321	CGACTTATAGAGGTGCGGAAGAGCTTGATGACTGTGAGGACAGACTTCAGCTCTCTCACG	1380
1543	ATGGAATACAGAAAGGAGAGAGAGAAATTGATGAGATTCGAAGGCTTAAACAGAAAAGA	1602
1381	ATGGAATACAGAAAGGAGAGAGAGAAATTGATGAGATTCGAAGGCTTAAACAGAAAAGA	1440
1603	GAGAGCTCATGTTTTCTTTTGCAGGAGGAGAACCGAAGATATGACTTTGCAAGAGCTGTCT	1662
1441	GAGAGCTCATGTTTTCTTTTGCAGGAGGAGAACCGAAGATATGACTTTGCAAGAGCTGTCT	1500
1663	GATCTAAGATATGGCGCAATTCAGAAGTGGATCTGCAATTTGCCCAACTTGAAGGAAC	1722
1501	GATCTAAGATATGGCGCAATTCAGAAGTGGATCTGCAATTTGCCCAACTTGAAGGAAC	1560
1723	TCTTCTGAAGAGAAATGTGATGCTCACAGAAAAAGTTTGGGCTTGAAACACATTTGCTGAGGTT	1782
1561	TCTTCTGAAGAGAAATGTGATGCTCACAGAAAAAGTTTGGGCTTGAAACACATTTGCTGAGGTT	1620
1783	GTGAGCCGTTGGAAGGGATTCAGTGAAGAGACTTTGGCCCAAAATGAGAAAGGAGGTTG	1842
1621	GTGAGCCGTTGGAAGGGATTCAGTGAAGAGACTTTGGCCCAAAATGAGAAAGGAGGTTG	1680
1843	ATTGCTCTGCTGATAGGTGTCATTAAGCGGTTGCGGACAGAAATCAAGCGGTAAATGCA	1902
1681	ATTGCTCTGCTGATAGGTGTCATTAAGCGGTTGCGGACAGAAATCAAGCGGTAAATGCA	1740
1903	GTTTCTGAGGCAATTTAAGGTCGAAGGCGAGGACTTGGTAGGCGCAACAGCCCACTGGGA	1962
1741	GTTTCTGAGGCAATTTAAGGTCGAAGGCGAGGACTTGGTAGGCGCAACAGCCCACTGGGA	1800
1963	TCATTTCTTATTCCTTGGACCAACTGTGTGTTGGCAAAACTGAGCTGCCAAGGCTCTTGCT	2022
1801	TCATTTCTTATTCCTTGGACCAACTGTGTGTTGGCAAAACTGAGCTGCCAAGGCTCTTGCT	1860
2023	GAGCAGCTGTTTGATGATGAAAACCTCTTAGTTTCCGATTTGATGTTCGGAATATATGGA	2082
1861	GAGCAGCTGTTTGATGATGAAAACCTCTTAGTTTCCGATTTGATGTTCGGAATATATGGA	1920
2083	CAACACTCTGTCTCTCGGCTCATTTGGGGGACACACAGGATATGTTGGTCAAGGAAAGGT	2142
1921	CAACACTCTGTCTCTCGGCTCATTTGGGGGACACACAGGATATGTTGGTCAAGGAAAGGT	1980
2143	GGCAACTAATCTGAGGCTGTGAGGAGGCGACCTTATTGTGTCTATCTCTTTGATGAAGTG	2202
1981	GGCAACTAATCTGAGGCTGTGAGGAGGCGACCTTATTGTGTCTATCTCTTTGATGAAGTG	2040
2203	GAGAAGGCTCATGTTGCTGTCTTCAACACTCTGCTCCAGTTTGGATGATGGTCCGATTG	2262
2041	GAGAAGGCTCATGTTGCTGTCTTCAACACTCTGCTCCAGTTTGGATGATGGTCCGATTG	2100
2263	ACAGACGGGCAAGGCGAGGACGTGATTTGAGNACTCGGTGTATATCATGACATCAAC	2322
2101	ACAGACGGGCAAGGCGAGGACGTGATTTGAGNACTCGGTGTATATCATGACATCAAC	2160

Zhu T, Glazov EA, Meins F, Wang X, Chang H;

823 GCACAAAGGATTGTGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTCG 882
661 GCACAAAGGATTGTGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTCG 720
883 TTGACATGGGTGCGTTAGTTGCTGCTCTAAATAACCGAGGAGAGTTTGAAGAAGGTTG 942
721 TTGACATGGGTGCGTTAGTTGCTGCTCTAAATAACCGAGGAGAGTTTGAAGAAGGTTG 780
943 AAATCTGTTTGAAGAAGATTGAGAGCCTGAAGGCAAGTGATCTCTTTATTTGATGAG 1002
781 AAATCTGTTTGAAGAAGATTGAGAGCCTGAAGGCAAGTGATCTCTTTATTTGATGAG 840
1003 ATTCAATTTGGTCTTGTGCTGCTGCAAACTGAAGGTCGATGGATGCACTTAATCTGTTT 1062
841 ATTCAATTTGGTCTTGTGCTGCTGCAAACTGAAGGTCGATGGATGCACTTAATCTGTTT 900
1063 AAGCCCATGTTAGCTTAGAGGGCAGCTTCGATGCAATTTGATGCTTAACGCTTTGAAGAATAC 1122
901 AAGCCCATGTTAGCTTAGAGGGCAGCTTCGATGCAATTTGATGCTTAACGCTTTGAAGAATAC 960
1123 AGGAATATGTTGAGAAAGATGCTGCCCTTTGAGAGGAGTTCCAAAGTCTATGTTGCG 1182
961 AGGAATATGTTGAGAAAGATGCTGCCCTTTGAGAGGAGTTCCAAAGTCTATGTTGCG 1020
1183 GAGCCAAAGTGTCCCTGACACCAATTAGTATCTCTTAGAGGACTCAAGGAGAGTATGAGGGA 1242
1021 GAGCCAAAGTGTCCCTGACACCAATTAGTATCTCTTAGAGGACTCAAGGAGAGTATGAGGGA 1080
1243 CATCATGTTGTCGAATCCAGACAGAGTCTTAAATGCTGCTCAGCTGTCTGCTCGT 1302
1081 CATCATGTTGTCGAATCCAGACAGAGTCTTAAATGCTGCTCAGCTGTCTGCTCGT 1140
1303 TACATAAATGTTGTCGCTTACCGGATTAAGCAATTTGATTTGTTGATGAGCTTTGTCG 1362
1141 TACATAAATGTTGTCGCTTACCGGATTAAGCAATTTGATTTGTTGATGAGCTTTGTCG 1200
1363 AATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATTAACCTTTGAAAGGAGG 1422
1201 AATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATTAACCTTTGAAAGGAGG 1260
1423 ATGCGAGCTGGAATTTGAATTTCAAGCTTTGAAAGGAGAGGATTAAGCCAGCAAGCT 1482
1261 ATGCGAGCTGGAATTTGAATTTCAAGCTTTGAAAGGAGAGGATTAAGCCAGCAAGCT 1320
1483 CGACTTATAGAGTCCGGAAGAGCTTGAATGATGCTGAGAGCAAGCTTCAAGCTCTCAG 1542
1321 CGACTTATAGAGTCCGGAAGAGCTTGAATGATGCTGAGAGCAAGCTTCAAGCTCTCAG 1380
1543 ATGAAATACAGAAAGGAGAGAGATTTGATGATTTGGAAGGCTTAAACAGAAAGA 1602
1381 ATGAAATACAGAAAGGAGAGAGATTTGATGATTTGGAAGGCTTAAACAGAAAGA 1440
1603 GAAGAGCTCATGTTTCTTTTTCAGAGGAGGAGAGGATGATGCTTGCAGAGCTGCT 1662
1441 GAAGAGCTCATGTTTCTTTTTCAGAGGAGGAGAGGATGATGCTTGCAGAGCTGCT 1500
1663 GATCTAAGATATGCGGCAATTTCAAGAGTGGATCTGCAATTTGCCCACTTCAAGGAACT 1722
1501 GATCTAAGATATGCGGCAATTTCAAGAGTGGATCTGCAATTTGCCCACTTCAAGGAACT 1560
1723 TCTTCTGAAGAGATGATGATCTCAAGAAAGCTTTGGGCTTGAACACATTTGCTGAGTT 1782
1561 TCTTCTGAAGAGATGATGATCTCAAGAAAGCTTTGGGCTTGAACACATTTGCTGAGTT 1620
1783 GTGAGCGTGTGAGAGGATTTCCAGTGAAGAGCTTTGGCCAAATGAGAGGAGAGGTTG 1842
1621 GTGAGCGTGTGAGAGGATTTCCAGTGAAGAGCTTTGGCCAAATGAGAGGAGAGGTTG 1680
1843 ATTGGTCTTCTGATAGTTGCAATAAGCGGTTTGTGGGACAGAAATCAAGCGGTAATGCA 1902
1681 ATTGGTCTTCTGATAGTTGCAATAAGCGGTTTGTGGGACAGAAATCAAGCGGTAATGCA 1740
1903 GTTCTGAGGCAATTTCTAAGGTCAGGGGAGGAGCTTTGGTAGGGGCAACAGGCCAAGTGA 1962

1741 GTTCTGAGGCAATTTCTAAGGTCAGGGCAGGACTTGGAAAGCCCAACAGCCAACTGGA 1800
1963 TCATCTTATTCCTTGGACCAACTGGTGTGGCAAAAATGAGCTGCCAGAGCTCTTGTCT 2022
1801 TCATCTTATTCCTTGGACCAACTGGTGTGGCAAAAATGAGCTGCCAGAGCTCTTGTCT 1860
2023 GAGCAGCTGTTTGAATGATGAAACCTCTTCTAGTTTCGATGATATGTCGGAATATATGGA 2082
1861 GAGCAGCTGTTTGAATGATGAAACCTCTTCTAGTTTCGATGATATGTCGGAATATATGGA 1920
2083 CAACACTCTCTCTCTCGCTCAATTTGGGCAACCAAGGATATGTTGGTCAAGAGAGT 2142
1921 CAACACTCTCTCTCTCGCTCAATTTGGGCAACCAAGGATATGTTGGTCAAGAGAGT 1980
2143 GGAACACTAATGAGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2202
1981 GGAACACTAATGAGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
2203 GAGAAGGCTCATGTTGCTCTCTCTCAACACTCTGCTCCAAAGTTTGGATGATGTCGATTTG 2262
2041 GAGAAGGCTCATGTTGCTCTCTCTCAACACTCTGCTCCAAAGTTTGGATGATGTCGATTTG 2100
2263 ACAGAGGGCAAGGCAAGGACAGTCTGATTTTCAGGACTCGGTCATATCATGATCAATCAAC 2322
2101 ACAGAGGGCAAGGCAAGGACAGTCTGATTTTCAGGACTCGGTCATATCATGATCAATCAAC 2160
2323 CTTTGGTGTGAAACCACTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2382
2161 CTTTGGTGTGAAACCACTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
2383 GACTGTGTGATGTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2442
2221 GACTGTGTGATGTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2280
2443 GAGATTGTGTGTGTTTCCAGCCCTTTTCCAGATGACCCAGTTCAGGAGGAGGAGGAGGAGG 2502
2281 GAGATTGTGTGTGTTTCCAGCCCTTTTCCAGATGACCCAGTTCAGGAGGAGGAGGAGGAGG 2340
2503 ATCAAGAGAGGCTTCTGCTCGGCTTGTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2562
2341 ATGAAAGAGGCTTCTGCTCGGCTTGTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
2563 GCTTTGGAATATATCTTTGTCAGAGAGGATTAAGTCCCGGTGATGTTGCTAGGCTATAAGG 2622
2401 GCTTTGGAATATATCTTTGTCAGAGAGGATTAAGTCCCGGTGATGTTGCTAGGCTATAAGG 2460
2623 AGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2682
2461 AGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2520
2683 GATGAAACCTCCACTGTTTATACATAGATGCAAGGCGCTGGTGATCTTGTGTACCGGGTAGAA 2742
2521 GATGAAACCTCCACTGTTTATACATAGATGCAAGGCGCTGGTGATCTTGTGTACCGGGTAGAA 2580
2743 AGTGGAGGCTTAGTGGAGCGTTTCAACAGGCAAGAGTCAAGTGTGCTGATTCATATTGCT 2802
2581 AGTGGAGGCTTAGTGGAGCGTTTCAACAGGCAAGAGTCAAGTGTGCTGATTCATATTGCT 2640
2803 AACGGGCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2862
2641 AACGGGCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2700
2863 GAAGATGAGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2898
2701 GAAGATGAGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2736

RESULT 4
AAI66066
ID AAI66066 standard; DNA; 6376 BP.
XX
AC AAI66066;

14-JAN-2002 (first entry)
Arabidopsis thaliana heat shock protein 101 gene (GenBank: AF218796).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
Arabidopsis thaliana; ds.

Arabidopsis thaliana.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51659.

Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (A166057-A166084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal,
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

Sequence 6376 BP; 1914 A; 1150 C; 1470 G; 1842 T; 0 other;

every Match 81.7%; Score 2535.8; DB 22; Length 6376;
st Local Similarity 85.7%; Pred. No. 0;
tches 3103; Conservative 0; Mismatches 2; Indels 516; Gaps 5;
1 AAAGTTATCAATTTTACAAATACCGCTATATCTGCTGATTTCTGCAAAAAGAGAA 60
866 AAAGTTATCAATTTTACAAATACCGCTATATCTGCTGATTTCTGCAAAAAGAGAA 925
61 GACTTTTACCAGAGAGAGTCTCTGGCTATTTGAGAGAACTCAACGAAACCAACCCAG 120
926 GACTTTTACCAGAGAGAGTCTCTGGCTATTTGAGAGAACTCAACGAAACCAACCCAG 985
121 TTCTCATATATCGTTTAA----- 138
986 TTCTCATATATCGTTTAAAGTAAATGATCGGACAACTCTGTTCTCATTTGTGTTT 1045
139 ----- 138
1046 TGTGTTTGTAGGTTTACAAAGATACCTGAGATTAGTTTCTTTTCTTTT 1105
139 --AGGATTACAAAGCTAATCGAAGTAAATCCAGAGAAATTCACACAGACAAACG 196
1106 TCAGGGATTACAAAGCTAATCGAAGTAAATCCAGAGAAATTCACACAGACAAACG 1165

QY	197	AGCAATTTGCTACAGCTCATGAGCTAGCTAGTGAATGCAGACATGCTCAATTCACCTCTT	256
DB	1166	AGCAATTTGCTACAGCTCATGAGCTAGCTAGTGAATGCAGACATGCTCAATTCACCTCTT	1225
QY	257	TGCAATTTAGCTGCTGCTTTGATCTCTGATCCACCGGTATATTTCTCAAGCAATCTCTA	316
DB	1226	TGCAATTTAGCTGCTGCTTTGATCTCTGATCCACCGGTATATTTCTCAAGCAATCTCTA	1285
QY	317	GTGCGGTGGGAGAGCGAGCTCAATCTCTGTAAGAGTGATCAATCAAGCCTTGAGA	376
DB	1286	GTGCGGTGGGAGAGCGAGCTCAATCTCTGTAAGAGTGATCAATCAAGCCTTGAGA	1345
QY	377	AGCTTCTCTTCAATCTCTCCCTCATGATATATCCAGCGAGTTCTAGTCTTTATTAAGG	436
DB	1346	AGCTTCTCTTCAATCTCTCCCTCATGATATATCCAGCGAGTTCTAGTCTTTATTAAGG	1405
QY	437	TCATTTCTGCTGCTCAAGCTGCTCAGAAATCAAGAGTGATCAATTTGGCTGTTGACC	496
DB	1406	TCATTTCTGCTGCTCAAGCTGCTCAGAAATCAAGAGTGATCAATTTGGCTGTTGACC	1465
QY	497	AGTTGATTATGGGTCTTCTTTGAAGATTCTCAATCAGGGATTGTTGAACGAAGTCGGT	556
DB	1466	AGTTGATTATGGGTCTTCTTTGAAGATTCTCAATCAGGGATTGTTGAACGAAGTCGGT	1525
QY	557	TAGCGACGGCGAGGTAAAGTCTGAGTTTGAAGAGCTTCGTGGGAAAGAGGGAAGAAAG	616
DB	1526	TAGCGACGGCGAGGTAAAGTCTGAGTTTGAAGAGCTTCGTGGGAAAGAGGGAAGAAAG	1585
QY	617	TTGAGAGTCTTTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTGG	676
DB	1586	TTGAGAGTCTTTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTGG	1645
QY	677	TTGAGCAAGCAGGAGAGCTTTCATCTGCTGATTCGTGATGAGGAGATTAGAGAGTCG	736
DB	1646	TTGAGCAAGCAGGAGAGCTTTCATCTGCTGATTCGTGATGAGGAGATTAGAGAGTCG	1705
QY	737	TGAGGATCTTTCAGGAGAGACGAAGAACTCTGCTGCTTATTTGAGAGCCAGGAGTTG	796
DB	1706	TGAGGATCTTTCAGGAGAGACGAAGAACTCTGCTGCTTATTTGAGAGCCAGGAGTTG	1765
QY	797	GTAAGACAGCTGCTGTTGAGGTTTACGACAAAGGATTGTGAAGAGAGATGTCCCAACA	856
DB	1766	GTAAGACAGCTGCTGTTGAGGTTTACGACAAAGGATTGTGAAGAGAGATGTCCCAACA	1825
QY	857	GTCTTACTGATGAGAGTTAATTTGCTTGGACATGGTGGCTAGTTGCTGCTGCTAAT	916
DB	1826	GTCTTACTGATGAGAGTTAATTTGCTTGGACATGGTGGCTAGTTGCTGCTGCTAAT	1885
QY	917	ACCGAGAGAGTTTGAAGAAAGGTTGAAATCTGCTTTTGAAGAGATTGAGGAGCTGAAG	976
DB	1886	ACCGAGAGAGTTTGAAGAAAGGTTGAAATCTGCTTTTGAAGAGATTGAGGAGCTGAAG	1945
QY	977	GCAAGTGATCTCTTTTATGATGAGATTCAATTTGGTCTTGGTCTGCGCAAACTGAAG	1036
DB	1946	GCAAGTGATCTCTTTTATGATGAGATTCAATTTGGTCTTGGTCTGCGCAAACTGAAG	2005
QY	1037	GCTCGATGATGAGCTAATCTCTTCAAGCCCATGTTAGTACAGGGCAGCTTCGATGCA	1096
DB	2006	GCTCGATGATGAGCTAATCTCTTCAAGCCCATGTTAGTACAGGGCAGCTTCGATGCA	2065
QY	1097	TTGGTGTCTCAACAGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTGGA	1156
DB	2066	TTGGTGTCTCAACAGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTGGA	2125
QY	1157	GGAGGTTCCAAAGTCTATGTTGCGGAGCCAAAGTGTGCTGACACATTTAGTATCTTGA	1216
DB	2126	GGAGGTTCCAAAGTCTATGTTGCGGAGCCAAAGTGTGCTGACACATTTAGTATCTTGA	2185
QY	1217	GAGGACTCAAGGAGAGTATGAGGACATCATGCTGTCGAATCCAAAGCAGAGCTCTTGA	1276
DB	2186	GAGGACTCAAGGAGAGTATGAGGACATCATGCTGTCGAATCCAAAGCAGAGCTCTTGA	2245

1277	TAAATGCTGCTCAGCTGTCTGCCTGGTTCATAAATCT	-----	1311
2246	TAAATGCTGCTCAGCTGTCTGCTCGTTACATAACTGGTATGTTAAGATTCTTAATCTCA	-----	2305
1312	-----	-G	1312
2306	GCTGATGTTTATGGTTTCATATAGTGTTGTTCTTGACATGACATTTTGTGGCGATTTAG	-----	2365
1313	GTCCGCATTTACC GGATAAAGCAAATGAATTTGGTTGATGAGCGTTGTGCGAATGTGAGAG	-----	1372
2366	GTCCGCATTTACC GGATAAAGCAAATGAATTTGGTTGATGAGCGTTGTGCGAATGTGAGAG	-----	2425
1373	TCCAGCTGTAGTCAACTCGAAGAGATTGATAAACCTTCGAAAGGAAGAGGATGCAGCTGG	-----	1432
2426	TCCAGCTGTAGTCAACTCGAAGAGATTGATAAACCTTCGAAAGGAAGAGGATGCAGCTGG	-----	2485
1433	AAATTTGAACCTTCAGGCCCTTGAAAAGGGAGAAGGATAAAGCCAGCAAAAGCTTCGACTTATAG	-----	1492
2486	AAATTTGAACCTTCAGGCCCTTGAAAAGGGAGAAGGATAAAGCCAGCAAAAGCTTCGACTTATAG	-----	2545
1493	-----	-----	1492
2546	AGGTATGATA CAGCTTTCTTTTCTCATCTAGTAGATAGCGCTTCAAAATATAAGACTATTTGT	-----	2605
1493	-----	-----	1492
2606	GCTTGTCTGGAAATCTTTGATTTGTAGCATCTGTTTTTCCGATGAGTCTTTGATATAATTT	-----	2665
1493	-----	-----	1532
2666	TTCTCAAACTATGAATCTGTAGSTGCGGAAAGAGCTTTGATGACCTTGAGAGA CAAGCTTCA	-----	2725
1533	GCCTCTCACGATGAAATACAGAAAGGAGAAGAGAGAA TTTGATGAGATTTCGAAAGCCTTAA	-----	1592
2726	GCCTCTCACGATGAAATACAGAAAGGAGAAGAGAGAA TTTGATGAGATTTCGAAAGCCTTAA	-----	2785
1593	ACAGAAAAAGAGAGAGCTCATGTTTTCTTGACGAGGCGAGAACGAAGATATGACCTTGC	-----	1652
2786	CACAGAAAAAGAGAGAGCTCATGTTTTCTTGACGAGGCGAGAACGAAGATATGACCTTGC	-----	2845
1653	AAGAGCTGCTCATCTAAGATATGCGCAATCTCAAGAAAGTGGAAATCTGCAATTTGCCCACT	-----	1712
2846	AAGAGCTGCTCATCTAAGATATGCGCAATCTCAAGAAAGTGGAAATCTGCAATTTGCCCACT	-----	2905
1713	TGAAGGAACTTTCTTCTGAAGAGAAATGTGATGCTCA CAGAAAAAGTTGGGCGCTGAACACAT	-----	1772
2906	TGAAGGAACTTTCTTCTGAAGAGAAATGTGATGCTCA CAGAAAAAGTTGGGCGCTGAACACAT	-----	2965
1773	TGCTCAGGTTGTCGAGCGTTGGA CAGGGAATTCAGTGA CAGAGACTTGGCCAAATGAGAA	-----	1832
2966	TGCTCAGGTTGTCGAGCGTTGGA CAGGGAATTCAGTGA CAGAGACTTGGCCAAATGAGAA	-----	3025
1833	GGAGAGTTGATTTGCTGTATAGTTTGCA TAAGCGGTTGTGGGACAGAAATCAAGC	-----	1892
3026	GGAGAGTTGATTTGCTGTATAGTTTGCA TAAGCGGTTGTGGGACAGAAATCAAGC	-----	3085
1893	GGTAATTCGAGTTTCTGAGGCAATTTTAAGGTC AAGGGCAGAGACTTGGTAGGGCAACA	-----	1952
3086	GGTAATTCGAGTTTCTGAGGCAATTTTAAGGTC AAGGGCAGAGACTTGGTAGGGCAACA	-----	3145
1953	GCACACTGGATCATTTCTTTCTTGACCAACTGTGTGTGCAAAACTGAGCTCGGCAA	-----	2012
3146	GCACACTGGATCATTTCTTTCTTGACCAACTGTGTGTGCAAAACTGAGCTCGGCAA	-----	3205
2013	GGCTCTTCTCAGCAGCTGTTTGATGATGAAAA CCTTTAGTTCGGATTGATATGTCGGA	-----	2072
3206	GGCTCTTCTCAGCAGCTGTTTGATGATGAAAA CCTTTAGTTCGGATTGATATGTCGGA	-----	3265
2073	ATATATGAA CAA CACTCTGTCTCTCGCCCTCATTTGGGCACACACAG	-----	2119
3266	ATATATGAA CAA CACTCTGTCTCTCGCCCTCATTTGGGCACACACACAG	-----	3325
2120	-----	-----	2119

3326	Db	TCTTTTAGTTTTCATATACCTTCGTGATTATGAATAATATTTTCTTACTTACACGATTACT	3388
2120	Qy	-----GGTATGTTGGTGCACGAGGAAGGTGCACAACATAACCTGAGCGCTGTGAGGAG	2168
3386	Db	ACACTTCTCAGGTAATGTTGGTTCAGGAGGAAGGTGCACAACATACTGAGCGCTGTGAGGAG	3445
2169	Qy	GCAGCCTTATTTGTTGTCATACCTCTTTTGATGAGTGGAAGGCTCATGTTGCTGTCTTCAA	2228
3446	Db	GCAGCCTTATTTGTTGTCATACCTCTTTTGATGAGTGGAAGGCTCATGTTGCTGTCTTCAA	3505
2229	Qy	CACCTGCTCCAAAGTTTTCGATGATCGTTCGATTGACAGACGGCAGGACGAGACAGTCGA	2288
3506	Db	CACCTGCTCCAAAGTTTTCGATGATCGTTCGATTGACAGACGGCAGGACGAGACAGTCGA	3565
2289	Qy	TTTTCAGGAACCTCGGTGATTAATCATATGACATCAAACCTTGGTCTGTAACAACCTCTTGACAGG	2348
3566	Db	TTTTCAGGAACCTCGGTGATTAATCATATGACATCAAACCTTGGTCTGTAACAACCTCTTGACAGG	3625
2349	Qy	GCTTAACTGGGAAGTTAAACAATGGAAGTGGCCCGGACCTGTGTGATGCGGAGGTGAGGAA	2408
3626	Db	GCTTAACTGGGAAGTTAAACAATGGAAGTGGCCCGGACCTGTGTGATGCGGAGGTGAGGAA	3685
2409	Qy	ACACTTCAGACACAGAGCTCTTGAAACAGGCTTGACGAGATTGTGGTGTTCGACACCCCTTTC	2468
3686	Db	ACACTTCAGACACAGAGCTCTTGAAACAGGCTTGACGAGATTGTGGTGTTCGACACCCCTTTC	3745
2469	Qy	ACATGACCAAGTTGAGGAAGTAGCTCGGCTTCAAATGAAAGAGTTGTGCTCGGCTTGC	2528
3746	Db	ACATGACCAAGTTGAGGAAGTAGCTCGGCTTCAAATGAAAGAGTTGTGCTCGGCTTGC	3805
2529	Qy	TGAAAGAGGAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGAGAG	2588
3806	Db	TGAAAGAGGAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGAGAG	3865
2589	Qy	TTATGACCC-----	2597
3866	Db	TTATGACCCGGTAAGTCTCTTCACATGATGCAGATGCATATCATTTGTTGATCTCTCAGG	3925
2598	Qy	-----GGTGTAT	2604
3926	Db	CTGGTTGTTCTCTTGACTCACTCTTGCTGGTTCCTTTCTTTGTTTGTGTTAGGTGTAT	3985
2605	Qy	GGTCTAGGCCTATATAGGAGATGGATGGAGAGAGAGGTGTCGACAGAACTGTGCAAGAATG	2664
3986	Db	GGTCTAGGCCTATATAGGAGATGGATGGAGAGAGAGGTGTCGACAGAACTGTGCAAGAATG	4045
2665	Qy	GTTGTGCGTGAGGAATTCGATGAAAACTCCACTGTTTTCATATAGATGCAGGCGCTGTGTAT	2724
4046	Db	GTTGTGCGTGAGGAATTCGATGAAAACTCCACTGTTTTCATATAGATGCAGGCGCTGTGTAT	4105
2725	Qy	CTTGTGTACCGGCTAGAAAGTGGAGTCTTAGTGGACGCTTCAACAGGCGAAGAGTCAGAT	2784
4106	Db	CTTGTGTACCGGCTAGAAAGTGGAGTCTTAGTGGACGCTTCAACAGGCGAAGAGTCAGAT	4165
2785	Qy	GTGCTGATTCAATTCGCTAAACGGGCCAAAGAGAAAGTATGATGCAGCTCAGGCGGTGAAGAG	2844
4166	Db	GTGCTGATTCAATTCGCTAAACGGGCCAAAGAGAAAGTATGATGCAGCTCAGGCGGTGAAGAG	4225
2845	Qy	ATCAGGATTCGAGGAATAGAAAGATGACGATAATGAGGAATGATCGAGGATTAAGTCTTT	2904
4226	Db	ATCAGGATTCGAGGAATAGAAAGATGACGATAATGAGGAATGATCGAGGATTAAGTCTTT	4285
2905	Qy	GACCTCCAAGTGTAGTTCCTTTTGAGGTTGTTGATGTATGATCTTGAGCTTTCATGCT	2964
4286	Db	GACCTCCAAGTGTAGTTCCTTTTGAGGTTGTTGATGTATGATCTTGAGCTTTCATGCT	4345
2965	Qy	CTGTTTTGGTTTTGACTCTTTTGGGTAGACTATAATGGTGTGAACTGTAAAAAAGTCTAC	3024
4346	Db	CTGTTTTGGTTTTGACTCTTTTGGGTAGACTATAATGGTGTGAACTGTAAAAAAGTCTAC	4405
3025	Qy	TTTATTATGTTTTCGTATATGTAATCATTTTTCATPAAGAAATCCATTGTACGTGCCATA	3084

4406 TTATTATGTTTCGTATATGTAATCAATTTATACATAAGAAATCCATTTGACGTCCCATTA 4465

3085 ATCTTTTGGGGTTTTTTCATA 3105

4466 ATCTTTTGGGGTTTTTTCATA 4486

BLT 5
6067

AAI66067 standard; DNA; 3049 BP.

AAI66067;

14-JAN-2002 (first entry)

Glycine max heat shock protein gene (GenBank: L35272).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
Arabidopsis thaliana; ds.

Glycine max.

W0200170329-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51662.

Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAI66071). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal,
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

Sequence 3049 BP; 850 A; 564 C; 863 G; 772 T; 0 other;

Query Match 52.8%; Score 1638.4; DB 22; Length 3049;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 2094; Conservative 0; Mismatches 646; Indels 18; Gaps 5;

161 AGATGAATCCAGAGAAATTCACACAGACAAACAGAGACAAATGCTACAGCTCATGAGC 220

134 AAATGAATCTCTGAGAGTTTACTCACAGACAAATGAAGCTCTTTTCGCTCGCCCGCAGC 193

221 TAGCTGTGAATGCAGGACATCTCAATTCATCTCTTTCATTTAGCTGGTGTGATCT 280

194 TCGCATGATGTTAGGCGCACCGCAATGACTCCATCCACTTGGCCCCCGCTGATT 253

281 CTGATCCACCGGTATATTTCTCAAGCAATCTCTAGTGC---CGTGGCGAGAACGAG 337
254 CGATCCCAACCGCATCTTCGTCTAGCGATAAACAAGCGCGCGCGGCGAGGAATCGG 313
338 CTCATCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAGCTTCCCTTCAAACTCTCTC 397
314 CACGCGCGGTGAGCGAGTGTGAAACAGGCTCTGAAGAAAGCTACCCCTGCCAGTCCCTC 373
398 CACCTGATGATATTTCCAGCGAGTTCCTAGTCTTATTAAGGTCAATTCGTCTCAAGCTG 457
374 CGCCGACGAGGTGCCGCGAGCACCACTCTGTAGGGCCATCAGGAGACACAGCGG 433
458 CTCAGAGCTCACGAGGTGATCTCATTTGGCTGTGTGACCACTGATTAATGAGGTCTTCTG 517
434 CGCAAAATCACTGGCGCACACGCGTTTGGCGGTGTGATCACTGATCTCTCGGAATCTCTG 493
518 AAGATCTCAATCAGGATTTGTTGAACGAAGTCCGTGTAGCGACGCGGAGGTTAAGT 577
494 AAGACTCCCAATCGGAGACCTGTGAAAGGAAGCGGGTTGCGGTGGCGAAGGTAGAT 553
578 CTGAGGTTGAGAAGCTTCGTGGGAAAGAGGAAAGTTTGAGAGTCTTTCAGGGACA 637
554 CGGAGTGTGATAGCTTCGTGGGAGGAGGAGGAGGTTGAGAGCGCTTCCGGGGATA 613
638 CAAATTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAGAGCTTG 697
614 CGAATTTCCAAAGCTTTGAAGACTTATGGCGGTGACCTTTGTTGAAACAGCGGGAGCTCG 673
698 ATCCTGTGATTTGGTCTGATGAGGAGATTTAGAGAGTCTGTCAGGATTTCTTCGAGGAA 757
674 ACCCTGTTATTTGGCGGTGACGAAGAGATTAGAGGTTGTGAGGATTTCTACCGGAGA 733
758 CGAAGACAATCTCTGTCTTATTTGAGAGCCAGGAGTTGGTTAAACAGCTGTGTTGAAG 817
734 CTAAGACAACCCCGGTTCTCGTTGAGAACCCGGGTGTGGGAAACCTGCGGTGTGGAAG 793
818 GTTTAGCAAAAGGATTTGAAAGAGATGTGCCCAACAGCTCTTATCTGATGTGAGATTAA 877
794 GGTTCGCACAGAGGATAGTAAGAGGGGATGTTCCAAGCAACCTTGTGATGTGAGGCTTA 853
878 TTTTCGTTGACATGGTGGCTAGTTCGTGCTGTCTTAATACCGAGGAGAGTTTGAAGAA 937
854 TTGCGTTGATATGGGGCGTGTGGTGGCGGTGCGCAAGTATAGAGCGGAGTTGAGGAGC 913
938 GGTTCGAATCTGTTTGAAGAGTGTGAGGACGCTGGAAGGCAAAAGTGAATCTCTTTATTG 997
914 GGTAAAGGCTGTTTGAAGAGTGTGAGGAGGCTGAGGGAGAGTGATCTCTTCAATTG 973
998 ATGAGATTCATTTGGTCTTGGTGTGCGCAAACTGAAGGCTCGATGATGAGCTAATC 1057
974 ATGAGATTCATTTGGTCTTGGTGTGCTGTGCTGTGAGAACTGGAAGGCTTCCATGATGCTGCTAATC 1033
1058 TCTTCAAGCCCATGTTAGTAGAGGCGAGCTTCGATGCTATGCTGTCTACAACTGAGTGAAG 1117
1034 TATTCAACCTATGCTTGTCTCGGCGCAGCTTAGGTGATTTGGTGGCCACACAGCTTGAAG 1093
1118 AATACAGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTTCCAACTGATG 1177
1094 AGTACAGGAGTATGTGGAAGAGATGCTGCATTCGAGAGGAGGTTCCCAACAGGTTTGTG 1153
1178 TTGCGGAGCAAGTGTGCTGACACCTTAGTATCTTCTTAGAGGACTCAAGGAGAGATG 1237
1154 TGGCGGAACTAGTGTGTTGATACCATTAGCATTTCTTGGTGGCTTGAAGAGAGATG 1213
1238 AGGACATCATGCTGTGCAATCCAAAGACAGAGCTCTTATAATGCTGCTCAGCTGCTG 1297
1214 AAGGCCATCACGGTGTGAGAAATTCAGGACCGGTGCTTTGGTTATGGAGCTCAATTGCTA 1273
1298 CTCGTTTACATACTCGTGGCATTTTACCGGATTAAGCAATTTGATTTGATGAGGCTT 1357
1274 ACCGGTATATATCTGGCGGTCTCTCTCTGACAGGCAATGATTTGTTGTCAGGAGCTT 1333

1358 GTGCAATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAACCTTTGAAAGGA 1417
1359 GTGCAATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAACCTTTGAAAGGA 1418
1360 GTGCAATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAACCTTTGAAAGGA 1393
1418 AGAGAGTGCAGCTGGAATTTGAATTTCAACCTTTGAAAGAGAGAGATTAAGCCAGCA 1477
1419 AGAGAGTGCAGCTGGAATTTGAATTTCAACCTTTGAAAGAGAGAGATTAAGCCAGCA 1478
1420 AGAGAGTGCAGCTGGAATTTGAATTTCAACCTTTGAAAGAGAGAGATTAAGCCAGCA 1453
1478 AAGCTCGACTTATAGAGTGGGAAAGAGCTTTGATGACCTGAGAGACAAGCTTCAGCCTC 1537
1479 AAGCTCGACTTATAGAGTGGGAAAGAGCTTTGATGACCTGAGAGACAAGCTTCAGCCTC 1538
1480 AAGCTCGACTTATAGAGTGGGAAAGAGCTTTGATGACCTGAGAGACAAGCTTCAGCCTC 1513
1538 TCAGATGAATAACAGAAAGAGAGAGAGATTTGATGATTTGAGAGCTTTGAGAGCTTT 1597
1539 TCAGATGAATAACAGAAAGAGAGAGAGATTTGATGATTTGAGAGCTTTGAGAGCTTT 1598
1540 TGAATGATGAAGTACCGAAAGAGAGAGAGAGCTTTGATGATTTGAGAGCTTTGAGAG 1573
1598 AAGAGAGAGCTCATGTTTCTTTGAGAGAGAGAGAGAGAGCTTTGATGATTTGAGAG 1657
1599 AAGAGAGAGCTCATGTTTCTTTGAGAGAGAGAGAGAGAGCTTTGATGATTTGAGAG 1658
1600 AAGAGAGAGCTCATGTTTCTTTGAGAGAGAGAGAGAGAGCTTTGATGATTTGAGAG 1633
1634 CTGCGAGCTTGAAGATGAGAGAGAGAGAGAGAGCTTTGATGATTTGAGAGAGAGAG 1717
1635 CTGCGAGCTTGAAGATGAGAGAGAGAGAGAGAGCTTTGATGATTTGAGAGAGAGAG 1693
1718 GAACCTTCTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1777
1719 GAACCTTCTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1778
1694 GGA---GCACTGAAGAGAGATCTGATGATGATGATGATGATGATGATGATGATGAT 1750
1778 AGTTTGTGAGCGCTTGGAGAGAGATTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1837
1779 AGTTTGTGAGCGCTTGGAGAGAGATTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1838
1751 AGTTTGTGAGCGCTTGGAGAGAGATTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1810
1838 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1897
1811 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1870
1898 ATGAGTGTGAGAGAGATTTGAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
1871 ATGCTGTGCTGAGGCTGCTGTTGATGATGATGATGATGATGATGATGATGATGATG 1930
1958 CTGGATCATCTTATTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017
1931 CTGGTTCATCTTATTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1990
2018 TTGCTGAGAGAGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2077
1991 TTGCTGAGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2050
2078 TGGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2137
2051 TGGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2110
2138 AAGTGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2197
2111 AAGTGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2170
2198 AAGTGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2257
2171 AAGTGGAG 2230
2258 GATGAG 2317
2231 GGTAACTGATGAG 2290
2318 CAAACCTGCTGCTGAG 2377
2291 CAAATCTTGGAG 2350
2378 CCGGAG 2437
2351 CTCGGAG 2410
2438 TTGAG 2497

Db 2411 TTGATGAATTTGCTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2470
QY 2498 TTCAATGAAG 2557
Db 2471 TACAAATGAAG 2530
QY 2558 ATGCTGCTTTTGGAG 2617
Db 2531 ATGAG 2590
QY 2618 TAAG 2677
Db 2591 TAAG 2650
QY 2678 AAATCCATGAAG 2731
Db 2651 AGATTGATGAG 2710
QY 2732 ACCGGGTGAG 2788
Db 2711 ACCGTGAG 2770
QY 2789 TGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2845
Db 2771 TGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2830
QY 2846 TGAAGATTCAG 2903
Db 2831 TGAAGATTCAG 2888

RESULT 6

AAI66068
ID AAI66068 standard; DNA; 3052 BP.
XX AAI66068;
XX
DT 14-JAN-2002 (first entry)
XX
DE Nicotiana tobacum 101 kDa heat shock protein gene (GenBank: AF083343).
XX
KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
KW Arabidopsis thaliana; ds.
XX
OS Nicotiana tobacum.
XX
PN W0200170929-A2.
XX
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-US08836.
XX
PR 20-MAR-2000; 2000US-150769P.
PR 18-APR-2000; 2000US-198116P.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Lindquist S, Queitsch C, Vierling E;
XX
DR WPI; 2001-639123/73.
DR P-PSDB; AAM51665.
XX

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has

increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3052 BP; 898 A; 536 C; 836 G; 782 T; 0 other;

Query Match 50.4%; Score 1564.4; DB 22; Length 3052;
Best Local Similarity 74.4%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 686; Indels 18; Gaps 5;
147 CAAAGCTAATCGAAGATGAATCCAGAGAAATTACACACAGACAAACGAGCAATTGC 206
85 CTATAGGTGACTATTATGAATCTCGAATAATTCACCCACAGACTAACGAGGCCCTTGC 144
207 TACAGCTCATGAGTACTGTGAATGAGAGACATGCTCAATTCATCTCTTTGCAATTAGC 266
145 TGGGGCACTCGAGTAGCACTATCCGAGGGGCATGCTCAATTTACGCCCTCTGCATATGC 204
267 TGGTCTTTCATCTCTCATCCACCGGTATATTTCTCAAGCAATCTTAGTCCCGGTGG 326
205 TGTGGCTTAAATCTGATCACATGGTATTTTTCGACAGCGATTTGTCATGCTGGTGG 264
327 ---CGAAGACGCGAGCTCAATCTGCTGAAGAGTGATCAATCAAGCCCTTGAAGAGCTTCC 383
265 GAATGAAGAGTAGTAAATTCAGTGGAGCGGTATTGAATCAAGCGATGAAGAAGTACC 324
384 TTCCATATCTCTCCACTCATGATATTCAGCGAGTTCAGTCTTATTAAGTCAATTGC 443
325 TTCTCAACACCGGCTCTGACGAAATCCCACTAGCACTTCCCTTATCAAGGTGTACG 384
444 TCGTCTCAAGCTGCTCAGAAATCAGAGGTGATCTCAATTTGGCTGTTGACCACTTGAT 503
385 CCGAGCAATCTGCGCAGAGTCTGCTGTGACAGCCATTTAGCAGTGGATCAGTTGAT 444
504 TATGGGTCTTTGAAGATTTCTCAATCAGGATTTTGTGAACGAAAGTGGTGTAGCGAC 563
445 TTTAGGACTCTAGAGACTCCCAATTTGAGATCTTTTGAAGAAGCTGGAGTGTAGTGC 504
564 GCGAGGTAAAGTCTGAGGTTGAGAGCTTCTGGGAAGAGAGGAGAAAGTTGAGAG 623
505 ATCAAGAGTGAATTCAGAGGTAGAGAACTTTAGAGGAAGAGGAAGAAAGTTCGAAG 564
624 TGCTTCAGGGGACACAAATTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCA 683
565 TGCTTCAGGGGACACCAATTCAGCACTCAACACTTATGGCGGTGATCTTTGGAACA 624
684 AGCAGGAAGCTTGATCTCTGATTTGCTGTGATGAGAGATTTAGAGAGTCTGTAGAGAT 743
625 AGCAGGAAGCTTGATCTCTGATTTGCTGTGATGAGAGATTTAGAGAGTCTGTGAGAT 684
744 TCTTTCGAGGAGACGAGAAACAACTCTGTGCTTATTTGGAGACCGAGAGTTGGTAAAC 803
685 TTTATCAAGGAGACTAAGAAACACCGGTTCTTATTTGGAGAGCCCGGTGTGGGTAAAC 744
804 AGCTGTGGTTGAAGTTTACACAAAGGATTTGMAAGAGATGTGCCCAACAGTCTTAC 863
745 AGCATGTTGMAAGGCTAGCAGAGAGATTTGACGTGTGATTTTCCAGATTAATTAGC 804
864 TGATGTGAGATTAATTTGTTGAGCATGGGTGCTGTAGTTGCTGGTCTTAAATVACGAGG 923
805 TGATGTAGGCTTATAGCATTTGGATATGGAGCGCTAGTTGCTGGAGCTAAAGTACAGAG 864
924 AGATTTGAAGAAAGGTTGAATCTGTTTGAAGAAGAGTTGAGGACGCTGAGGCAAGT 983
865 TGAATTTGAAGAGAGGCTGAAGGCTGTGCTGAAAGAGTGTGAAGAAGCGGAGGAAAGT 924

QY 984 GATTCTCTTTATTGATGAGATTTCATTTGGTCTCTGCTGCAAAAACCTGAAGGCTCGAT 1043
DB 925 AATATCTTTTCATTGAGAGATACATTTAGTCTCGTCTGCTGGACAGAGAGGCTCAT 984
QY 1044 GGATCAGCTAATCTGTTCAAGCCCATGTTAGCTAGAGGGCAGCTTCGATGCAATTTGGTC 1103
DB 985 GGATCTGCTAATCTGTTTAAAGCCAATCTAGCCAGAGGTCATTTACGGTGCATTTGGTGC 1044
QY 1104 TACAACGCTTGAGAGATACAGGAATATGTTGAGAGAGATGCTGCTTTGAGAGAGGTT 1163
DB 1045 AACTACACTCGAGGAGTACAGGAAGTATGTTGAGAGGATGCTGCATTTGAGAGCGGTT 1104
QY 1164 CCAACAGTCTATGTTGCGAGCCCAAGTGTGCTGACACCATTTAGTATCTTTAGAGGACT 1223
DB 1105 CCAGCAGGTGATGTTGCTGAGCCTAGTGTGCTGACACTATTAGTATCTCCGTTGGGT 1164
QY 1224 CAAAGAGAGTATGAGGACATCATGTTGTGCGAATCAAGACAGAGCTCTTATAAATGTC 1283
DB 1165 GAAGGAGAGGTATGAAGGCGCATCATGTTGTCAAAATTCAGGACAGAGCTCTTTGTAGTGC 1224
QY 1284 TGCTCAGCTGTCTGCTGCTTACATTAACCTGCTGCGCATTTACGGATTAAGCAATTTGTTT 1343
DB 1225 TGCCAGCTCTCATCTCGGTACATTTACAGGTCGACATCTGCCAGATAGGCTTATTTGACT 1284
QY 1344 GGTTCATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGATAGTCAACCTGAAAGAGATTGA 1403
DB 1285 AGTTGATGAGCTTGTGCAATTTGAGAGTTCAGCTTGATAGTCACTCTGAGGAAATTTGA 1344
QY 1404 TAACTTTGAAGAGAGAGATGAGCTGGAATTTGAATTTCAAGCTTTGGAAGAGGAGAA 1463
DB 1345 CAATCTTGAGAGGAGAGAAATTCAGCTAGAGGTTGAATTTCAAGCTCTCGAGAGAGAAA 1404
QY 1464 GGATTAAGCCAGCAAGCTCGACTTATAGAGTCCGGAAGAGCTTGATGACCTGAGAGA 1523
DB 1405 AGCAAGCTAGCAAGCAGCTCTCATAGAGTGAAGAGAACTTGATGATTTGAGAGA 1464
QY 1524 CAAAGCTTCAGCTCTCAGACTGAATAACAGAAAGAGAGAGAAATTTGATGAGATTTCG 1583
DB 1465 CAACTCCAACTTTGATGATGAGGTACAGAGAGAGAGAGAGATAGATGAGTCTCG 1524
QY 1584 AAGGCTTAAACAGAAAGAGAGAGCTCATGTTTCTTTGAGAGGAGAGAGAGAGAGATA 1643
DB 1525 CAGGCTCAAGCAAAAGCGGATGAGCTCATCTATGCTTTTCAAGAGCTGAAAGAGAGATA 1584
QY 1644 TGACCTTCAAGAGCTGCTGATCTAAGATATGGGCAATTTCAAGAGTGGATCTGCAAT 1703
DB 1585 TGATCTGGCGAGGCGAGAGATCTGAGATATGGGCAATTTCAAGAGTGGAACTGCAAT 1644
QY 1704 TGCCCAACTTGAAGAACTTTCTTCTGAGAGAGATGCTGATGCTCAGAGAAAACGTTGGGCC 1763
DB 1645 AGCAATCTTGAGAGTACTTCAGCTGAA--AGTACAATGCTAACAGAGACTGTGGGTCC 1701
QY 1764 TGACACATCTGCTGAGGTTGTGAGCGGTTGACAGAGGATTTCCAGTACAGAGACTTGGCCA 1823
DB 1702 TGATCAGATTGCGAAGTTGTGAGTCTGCTGAGTGGTATTTCCGGTCTCAAGGCTTGGTCA 1761
QY 1824 AAATCAGAGAGAGGTTGATTTGTTGCTGATAGTGTGATGAGCGGTTGTTGGGACA 1883
DB 1762 GAATCAGAGAGAGAACTGATTTGGTCTTTGGCGATAGATTGCCAAAGAGTGGTTCGGGCA 1821
QY 1884 GAATCAGCGGTAAATGACAGTTTCTGAGGCAATTTCTAAGGTCAAGGGCAGGACTTGGTAG 1943
DB 1822 AGATCATGAGTTAGAGCTGTTGCTGAAGCGCTGTTAAGGTCCAGAGCTGGTTTAGGAG 1881
QY 1944 GGCACAAAGCAAGCTGATCATTTCTTATCTTTGAGCCAACTCGGTGTGGCAAACTGA 2003
DB 1882 GCCACAGCAAACTGTTTCAATCTCTTTCTTTGGGCGCAACTGCTGTGTGAAAGACAGA 1941
QY 2004 GCTCGCCAGAGCTCTGCTGAGCAGCTGTTTGTGATGATGAACCTCTTAGTTCCGATTGA 2063
DB 1942 GCTCGCTAAAGCTTTGACAGAGAGCTCTTTGATGATGAATAACTGATGATCAGATAGA 2001

Query Match	43.3%;	Score 1345;	DB 22;	Length 3084;
Best Local Similarity	69.1%;	Prod. No. 0;		
Matches 1904;	Conservative 0;	Mismatches 835;	Indels 18;	Gaps 4;
QY	141	GGATTACAAAAGCTAAATCGGAAGATGAATCCAGAGAAATTACACACAAGACAACACGAGAC	200	
DB	123	GGAGTAGCAAGAGAGACGCCCATGAATCGGACCACTTACCACAAGACGACGAGGC	182	
QY	201	AATTGCTACAGCTCATGAGCTAGCTGTGTAATGCAGACGATGCTCAATTCACTTCCTTTGCA	260	
DB	183	GATCTGGGGGCGCACGAGATTTCGGTGGAGGCGCGCCACGCGCAGCTACGCGCGTGCA	242	
QY	261	TTTACTGCTGCTTTGATCTCTGATCCCAACCGGTATATTTCCTCAAGCAAACTCTCTAGTGC	320	
DB	243	CCTGGCCGCAGTGTCTGGCTGGGACAAAGGGCGGCATCTCTCGGACGAGCCATCACGGGGC	302	
QY	321	CGGTGGCGAGAACCGAGC-----TCATCTGCTGAAGAGTGATCAATCAAGCCTTGAA	374	
DB	303	GTCCGGGGGCGACGAGGCGCGCGGGACTCGTTTCGAGCGCGTGTGAAACACTCGCTCAA	362	
QY	375	GAAAGTTCCTTTTCAAACTCTCTCCACTGATGATATTCCAGCGAGTTCTAGTCTTATTAA	434	
DB	363	GAAAGTCCGCTGCGAGTCCCGCGCGCGAGCTCCGTCCTCGGCGTCCACGGCACTGATCAA	422	
QY	435	GGTCAATTGCTCGTCTCAAGCTGCTCAGAAAGTCACGAGGTGATATCATATTTGGCTGTGGA	494	
DB	423	GGTCACTCCGCGGGCGAGTCCGCGCGAGAAAACCGCGGGGACTCTCACTTCGCGCGTGA	482	
QY	495	CCAGTTGATTAATGGTCTCTTTGAAGATTCTCAAACTCAGGGAATTTGTTGAACGAAGTCGG	554	
DB	483	CCAGTCTGCTCGGCTGCTCGAGGACTCGCAGATCTCCGACTGCCTCAAGGAGCGCGG	542	
QY	555	TGTAGCGACGCGAGGGTAAAGTCTGAGTTTGAGAAGTTCGTGGGAAGAGTGGGAAGAA	614	

Zea mays.

543	CGTGTCCGCGCGCGCGGTGCGCGCGAGCTTGAGAAAGCTCCGCGCGGAGGCGCGCG	QY	1695	ATCTGCAATTCGCCAACTTGAGGAACCTTCTTCTGAAGAGAACTGTGATGTCTCACAGAAA	1754
615	AGTTGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTTAAAGACTTATGGAAGAGATT	Db	1683	CGCGCGGATCTTCCAGGTGGA--GAGCGAAACAGGGGAGAACTGTATGTCTCACCGAAAC	1739
603	CGTGGAGTCCGCGTCCGCGGACACCAACTTCCAGGCGCTCAAGACATACGCGCGGACCT	QY	1755	CGTTGGGCTCGAAACACATTCGAGGTTGTGAGCGTTGACAGGGATTCAGTGACGAG	1814
675	GGTTGAGCAAGCAGGAGAGCTTGATCTCTGTGATGTGTCTGATGAGGAGATTAGAGAGT	Db	1740	CGTCCGCGCTGAAACAAATTCAGAGGTGGTGGCGGTTTGGACGGGTATTCAGTGACCGG	1799
663	CGTTGAGCAGCGCGGAGCTGACCCCGCTCATCGGCGCGAGAGAGATCCGCGCGT	QY	1815	ACTTGGCCCAATCAGAAAGAGAGGTTGATTTGGTCTTCTGCTGATAGGTTGCAAGCGGT	1874
735	CGTGAGATTTCTTCCAGGAGAACGAAACAACTCTGTCTTATTCGAGAGCAGGAGT	Db	1800	GCTTGGCCAGAACGACAGGAGAGGCTTGGCTCGGCTTGGCTGACAGGCTTCCACAGAGGT	1859
723	GGTGGCATCTCTCGCGCGGACATAGAACCAACCCCGTCTCATCGGAGCGCGCGT	QY	1875	TGTGGGACAGAAATCAGGGTAAATTCAGATTTCTGAGGCAATTCAGGCTCAAGGGCAGG	1934
795	TGTTAAACACAGCTGTGTGTTGAGGTTTATGACAAAGGATTTGTAAGAGAGATGTGCCAA	Db	1860	GGTCCGCGCAGACAGAGGCTGTGAGCGCGCTCGCAGAGAGCGGTGTGAGGTCGAGGGCGG	1919
783	TGGCAAGACGCGCGTCTGTGAGGCGCTCGCGAGCGCATGTTTCCGCGCGAGCGTGC	QY	1935	ACTTGGTAGGSCACAAACAGCCAACTGGATCATTTCTTATTCCTTGGACCACTGTGTG	1994
855	CAGTCTTACTGATGTGAGATTATTTCTGTGGACATGGGTGCGTTAGTTGCTGTGCTAA	Db	1920	TCTTGGCAGGCCAACACAGCCCACTGGCTGCTCTGTTCTGTGGTCCGACTGGCGTGGG	1979
843	TAACCTCTCTCGACGTCGCGCTCATCGCGCTCGACATGGCGCGCTCGTCCGCGGCGCAA	QY	1995	CAAACTGAGCTCGCCCAAGGCTTCTGTGAGCAGCTGTGTTGATGATGAAACCTCTTGT	2054
915	ATACCGAGGAGGTTTGAAGAAAGTTGAAATCTGTTTGAAGAAAGTTGAGACGCTGA	Db	1980	GAAACTGAGCTAGSCCAAGGCCCTAGCCGACAGCTGTTCGACGACGAGAACTTCTTGT	2039
903	GTACCGCGCGAGTTCAGAGAGCGGCTCAAGCCGCTGCTCAAGGAGTGGAGAGCGCA	QY	2055	TCGATTTGATATGTCGGAATATATGGAACAACTCTGTCTCTGCTCTCATTTGGGGCACC	2114
975	AGCAAGTGAATCTCTTTATGATGAGATTCAATTTGTTGTTGTTGTTGTTGTTGTTGTTG	Db	2040	CCGATCTGACATGTCCGAGTACATGAGAGCAGCTCGGTTGCTCGCTCATTCGAGAGACC	2099
963	GGGGAAGTCAATCTCTTCATCGACAGATACACCTCGTCTCGCGCGCGGACAGCGA	QY	2115	ACCAGGTTATCTGCTCACAGGAAAGTGGACAACTAACTGAGGCTGTGAGGAGCGGACC	2174
1035	AGGGTCGATGAGTGAATCTGTTTCAAGCCCATGTTAGCTAGAGGAGCAGCTTCGATG	Db	2100	GCTTGGCTACGTCGCGCATGAGAGGTTGGGAGCTGACTGAAACAGTGGAGGAGGCC	2159
1023	GGGTTCCATGGACGCGGCCAACCTGTTCAAGCCCAATGCTGGCGAGGGGACAGCTCAGGTG	QY	2175	TTATTTGTCTACTCTTTGATGAGTGGAGAGGCTCATGTTGCTGTCTTCAACACTCT	2234
1095	CATTGTGCTCAACAGCTTGAAGATACAGGAATATGTTGAAAGATGCTGCTCTTGA	Db	2160	GTACAGCGTGTCTGTTTCGACGAGGTCGAGAGGCCCATGTCGCCGTGTTCAACACCTT	2219
1083	CATCGCGCCACACAGCTGAGAGGTTACCGCAATGACTGAGAGGAGCGAGCGTTGCA	QY	2235	GCTCCAGTTTGTGATGATGATGATGAGAGAGGCTGATGATGATGATGATGATGATGATG	2294
1155	GAGAGGTTCCAAACAGTCTATGTTGGGAGCCAGTGTGCTGACACCAATTAGTATCT	Db	2220	GCTCCAGGTCCTCGACGAGCGGAGTTGACGATGAGGAGGAGGAGGAGGAGGAGGAGG	2279
1143	CGCGCGGTTCCAGCAGGTGTTGCTCGCGGAGCGGAGGCTGCGCGACACCGTCAGCATCT	QY	2295	GAACTCGGTGATATCATGATCAACCTTGGTGTGAAACACCTCTCTGAGGAGGCTAAC	2354
1215	TAGAGGACTCAAGGAGATGATGAGGACATCATGTTGTCGATGTCGATGTCGATGTCGATG	Db	2280	GAAACAGCTGATCATGATCATGATCATGATCATGATCATGATCATGATCATGATCATG	2339
1203	GAGGGAGCTCAAGGAGATGACAGGGGACACCTGCTGAGGATCCAGGAGCGGCGCT	QY	2355	TGGGAAGTAAACATGGAAGTGGCCCGGAGCTGTGTGATGCGGAGGAGGAGGAGGAGG	2414
1275	TATAATGCTGCTCAGCTGCTCTGTTACATAAACTGGTGGGCTTTACCGGATAAAGC	Db	2340	GGGCAAGAACTCCATGAAGGTCGCTCGCATCTGCTCATGACGAGGAGGAGGAGGAGG	2399
1263	CGTGGTCGCGCACAGCTATCCGCGAGGTACATCATGGGTGCGCACCTGCTGACAAAGC	QY	2415	CAGACAGAGCTCTTGAACAGGCTTGAAGAGGTTGAGGATGAGGATGAGGATGAGGATG	2474
1335	AATTGATTTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGTATGATCACTGA	Db	2400	CGGACCGGAGCTGCTGAAACCGCTTCGACGAGATCGTATCTTCTGCTCTCTCTCTCT	2459
1323	CATAGACCTGGTGGAGAGGCTTGGCGCAATGTGAGGAGTCCAGCTCGACAGCGCGGA	QY	2475	CCAGTTCAGGAAAGTAGCTCGGCTTCAAAATGAAAGAGCTTGTGCTCGGCTTGTGAAAG	2534
1395	AGAGATTGATTAACCTTGAAGGAGGAGGATGAGCTGGAATTTGAATTTCAAGCTTGA	Db	2460	GCAGCTGAGGAGGAGTGGCTCGCTTCAGATGAGGATGAGGATGAGGATGAGGATGAG	2519
1383	GGAGATTGATTAACCTGAGAGGAGGAGGATCCAGCTTGAAGTTGAGTTGAGCTTCA	QY	2535	AGGAGTTGCTTTGGCAGTCACTGATGCTGTGTTGGAATATATCTTGGCAGAGGATGAG	2594
1455	AAGGAGAGGATTAAGCCAGCAAGCTTCACTTATAGAGGTCGGAAGAGGATGATGATGA	Db	2520	GGGATCTGCTTGGCTGTGACCGACCGCGCATGGAATCATCTTGTCTCTCTCTCTCTCT	2579
1443	GAAGGAGAGGACAGCCAGCAAGCCCGCTGATCGAGTTCAGGATTCAGGAGGATTTGA	QY	2595	CCCGGTGATGCTAGGCTTAAAGGATGATGAGGAGGATGAGGAGGAGGAGGAGGAGG	2654
1515	CTTGAGAGACAGCTTCCAGCTTCAAGGATGAAATACAGAAAGGAGGAGGAGGAGGAGG	Db	2580	TCCCGTGTACGCGCGCGGCGCAATCAGAGGATGATCAGAGAGGAGGAGGAGGAGG	2639
1503	TCTGAGGAGCAAGCTGAGCGCTTCAAGGATGAGGATGAGGATGAGGATGAGGATGAG	QY	2655	GTCAAAGATGCTGTGCTGAGGAAATCGGATGAAATCTCCACTGTTTACATAGATG---	2710
1575	TGAGATTCGAGGCTTAAACAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG	Db	2640	GTGAGAGATGCTGATCCAGGAGGAGATCGACGAACTGCAAGCTTACATCGAGCGCG	2699
1563	TGAGATCAGGAGCTGAAGCAGCGCGCGGAGGAGCTTCAAGTTTCAAGGAGGAGGAGG	QY	2711	--CAGGCGCTGCTGATTTGTGATCCCGGTAGA---AAGTGGAGGCTTGTGAGAGGCTTC	2765
1635	ACGAAGATATGACCTTCAAGAGCTGCTGATCTAAGATATGCGGCAATTAAGAGTGA	Db	2700	GCCCGGCAAGGACGAGCTGCTACAGGTTGACCGGAGCGGCGGCTGCTGAGGAGGAGG	2759
1623	CGCGCGGATGGAATCTGCGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG				

223 GCTGTGAATGCAGGACATGCTCAATTCACTCCTTTGCAATTAGCTGGTGCCTTGATCTCT 282

```
1234 TCAGGTCATCATGCGTCCGCAATTTGCGCTGATAAAGCAATCGATTGGTTGTCAGAGGCC 1293
1357 TGTGCAATGTGAGAGTCCAGCTTGATGTAGTCTACCTGAGAGGATTCATTAACCTTGAAGG 1416
1294 TGGCAAAATGTGAGGGTGAACATTGACAGCCAGCTTGAGAGATGACAACTTCGAGAGG 1353
1417 AAGAGATGAGCTGGAAATTTGAATTCAGCCCTTGGAAGGAGAGAGATAAAGCCAGC 1476
1354 AAGAGATCCAGTTGGAAGTCGAGCTCCATCGCTCGAGAAGGAGAGCAAGGCTAGC 1413
1477 AAGCTCGACTTATAGAGGTGCGGAAGAGCTTGATGACTGAGAGACAGCTTCAAGCT 1536
1414 AAGCTCGGCTAGTTGATGTGAGAAAGAGTTGGACGATCTGAGACACAGCTGACGCG 1473
1537 CTCACGATGAATACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGGCTTAAACAG 1596
1474 CTCGATGAGTATTCGAAGGAGAAAGAGAGATCGACGAGATCAGAGAGCTGAGAGCAG 1533
1597 AAAAGAGAGAGCTCATGTTTCTTTTCAGAGAGGAGAGAGAGATATGACCTTGCAGA 1656
1534 CGCGCGAAGAGCTGCAGTTCACTCTGACAGAGGCGGAGCGCGGATGATTTGGCCGCG 1593
1657 GCTGCTGATCTAGATATGCGCAATTTCAAGAGTGGATCTGCAATTTGCCCACTTGA 1716
1594 GTGCTGATCTCAGATACGCTGCTCTGACAGAGGTTGACGCTGCCAATTTGCTAAGCTGGAG 1653
1717 GGAATCTCTTGAAGAGAGATGTGATGCTCACAAGAAACGTTGGGCTGAAACACATTTGCT 1776
1654 GG---TGAGACCGCGGAGAAATTTGATGTTTAAAGAGACTGTGCGCCGACGAGATTGCT 1710
1777 GAGGTTGTGAGCCGTTGGACAGGATTTCCAGTACGAGACTTGGCCCAAAATGAGAGAGG 1836
1711 GAGTGTGAGCCGCTGGACTGCTATTTCTGTACCAGGCTCGGACAGAAATGAGAGGCG 1770
1837 AGTTGATTTGCTTCTGCTGATAGTTGATAGCGGTTGTGGGACAGATCAAGCGGTA 1896
1771 AGCTGATCGGGTGGCAGATCACTGCTCATCGAGGGTGTGGACAGATGAGGCGGTC 1830
1897 AATGCACTTTCTGAGGCAATTTCTAAGTCAAGGCGCAGGACTTGTGAGGCGACAAAGCA 1956
1831 AATGCACTTGGAGAGGCTTTCTACGTTGAGGGCTGGCTTGGACGCGCGAGAGCT 1890
1957 ACTGATCATTTCTTATTCCTTGGACCAACTGTTGTTGGCAAACTGAGCTGCGCAAGCT 2016
1891 ACTGTTCTATTTCTTCTTGGACCGACCGGTTGCGAAACACGAGCTCGCAAGGCT 1950
2017 CTGCTGAGAGCTGTTGATGATGAATAAATCTTAGTTCGAGTTGATGATGCGGATAT 2076
1951 CTAGCTGAGAGCTGTTTGTATGACGAGAACTCTGCTGCTGCGCATCGACATGCTGAATAC 2010
2077 ATGCAACAACTCTGCTCTCGCTCATTTGGGCGACACCGAGGATGTTGTTGTCACGAG 2136
2011 ATGAGCAGACTTCGTTGCGCGCTTAATCGAGGCCACCTCGATGATGTTGTCATGAA 2070
2137 GAGGTGGAACAATACTGAGGCTGTGAGAGGAGGACCTTATTTGTGTCTATCTTTGAT 2196
2071 GAAGGGGGGAGCTGACCGAGCAAGTGAAGAGGAGGACATACAGTGTCTATCTTTGAC 2130
2197 GAGTGAAGAGGCTCATGTTGCTGCTTCAACACTGCTGCTCAAGTTTGGATGATGTT 2256
2131 GAGTTGAGAGGCGCATGTGGCGGTTTCAACACTGCTGCTGAGGCTGCGAGCATGGG 2190
2257 CGATTGACAGCGCGGAGGAGGAGGAGGAGTCTGATTTTCAGGAACCTCGGTGATATCATGACA 2316
2191 CGTTGACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2250
2317 TCAAACTTGTGCTGAACACCTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2376
2251 TCAAACTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2307
2377 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2436
2308 GCTGCTGATCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2367
```

2437 CTTGACGAGATGTTGGTGTTCGACCCCTTTTCATCATGACCAAGTTGAGGAAAGTAGCTCGG 2496

2368 CTGACGAGATCGTCACTTCCTTCGATGATGACAGCTGCGGAGGTCGCTCGG 2427

2497 CTTCAAAATGAAGAGCTTGTGTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2556

2428 CTTCAAGATGAAGATGTGGCAGTCCGCTTTCGCGAGAGGGGCGTTGCTCTGCGCGTCAAC 2487

2557 GATGCTGCTTTGGACTATCTTTCGCGAGAGATTTATGACCGGCTGATGCTGCTGCTGCTGCTGCT 2616

2488 GACGCGCGCTTGGAGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2547

2617 ATAAGGAGATCGATGGAAGAGAGAGTGTGTCAGAGAACTGTCAAAAGATGCTTGTGCTGCTGCTGCT 2676

2548 ATCCGAGATCGATGGAAGAGAGAGTGTGTCAGAGAACTGTCAAAAGATGCTTGTGCTGCTGCTGCT 2607

2677 GAATCGATGAAATCTCCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2730

2608 GAGATCGAGAGAACTCCACGCTGTACATGACGCTGCGCCGACCAAGGAGGAGCTGAC 2667

2731 TACCGGCTAGAAA---GTGAGGCTCTAGTGCAGCTTCAACAGGCAAGAGTCAAGTGTG 2787

2668 TATGCGCTGCAAGACGAGAGGCTGTGTAACGCGCACCGGCGCAAGTCCGACATC 2727

2788 CTGATTCATATTGCTTAAACGCGCCAAAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2847

2728 CTGATCCAGGTTCTTAGCGAGCTGTTGGGGGCGATGCGGCGCACGCGCTGAAGAGATG 2787

2848 AGGATCGAGGA 2858

2788 AAGATCATGCA 2798

RESULT 9

AAI66075

ID AAI66075 standard; DNA; 2821 BP.

XX AAI66075;

XX 14-JAN-2002 (first entry)

XX DE *Triticum aestivum* 10kDa heat shock protein gene (GenBank: AF083344).

XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

XX canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

XX *Arabidopsis thaliana*; ds.

XX *Triticum aestivum*.

XX WO200170929-A2.

XX 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US08836.

XX 20-MAR-2000; 2000US-190769P.

XX 18-APR-2000; 2000US-198116P.

XX (ARCH-) ARCH DEV CORP.

XX Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.

XX P-PSDB; AAM51661.

XX Transgenic plants with improved heat stress tolerance, useful for

XX producing animal feed, oil and synthetic products -

XX Claim 4; Page -; 91pp; English.

XX The invention relates to a transgenic plant, comprising a genetic

XX construct comprising a promoter operatively linked to a nucleic acid

sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAI66057-AAI66084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 2821 BP; 524 A; 902 C; 1035 G; 360 T; 0 other;

Query Match 39.8%; Score 1236.2; DB 22; Length 2821;
est Local Similarity 67.3%; Pred. No. 0;
Matches 1847; Conservative 0; Mismatches 869; Indels 30; Gaps 6;
156 ATCGAGATGATCCAGAGAAATTCACACACAGACAAAGAGACAAATGCTCAGACTCA 215
58 AGCGCCCATGAAACCCGGAACATTCACGCAACAGACCAACAGAGCGCTGTGGCGGCGCA 117
216 TGAGCTAGCTGTGAATGACGAGCATGCTCAATCTCACTCTTTTGCATTTAGCTGTGCTTT 275
118 CGAGCGCGCTCGAGGCGCGCCACGCGCAGATCACGCGTTTGCACCTGGCGGCGGCT 177
276 GATCTGTATCCACCGGTATATTTCTCAGCAATCTCTAGTCCGCGTGGCGGAGAA--- 332
178 GCGCGCGGACAAAGTCGGGCACTCTCGCGCGCGCGCTCGCGGGCGCTCCGCGGGAATGC 237
333 ---CGCAGCTCAATCTGCTGAAGAGTGAATCAATCAAGCTTTGAAGAGCTTCTCTTCA 389
238 GTCCCGGGGAGTCTGTCGAGCGGCTGCTCGCGCGCGCTCAGGAAGCTGCGCTCGCA 297
390 ATCTCTCCACTGATGATATTTCCAGCGAGTTCTAGTCTTATTAAGGTCTATTCGTGTGC 449
298 GTCCCGCGCGGACTCGTCTCGCGCTCCACGCGCTCATCAAGGCTATCCGCGCGC 357
450 TCAAGCTCTCAGAGTCAAGAGTGAATCTATTTGGCTGTGACCAAGTGAATATGCG 509
358 GCAGTCGCGCGAGAGAGCGCGGGGACTCGCACCTCGCGCTCGACCAAGTCTCATGG 417
510 TCTTCTGAAGATTTCTCAATCAGGATTTGTTGAACGAGTCGCTGAGCGAGCGCGAG 569
418 CTCTCTCGAGAGCGCGAGATCGCGAGTCTCAAGGAGCGCGCTGCTCGCTCGG 477
570 GGTAAAGTCTGAGGTTGAGAGCTTGT---GGGAAAGAGGGAAGAGTTGAGAGTGC 626
478 GGTGCGCGCGAGCTCGACAGCTCGCGCGGGGACAACTCGCGCAAGGTCGAGTCCGC 537
627 TTCAGGGACACAAATTTCAAGCTTTAAGACTTATGAAGAGATTTGTTGACCAAGC 686
538 CTTGCGGACACCACTTCCAGCGCTCAAGAGTACGCGCCGCGACCTCTGTCAGGTGCG 597
687 AGGGAAGCTTGTATCTCTGTGATGAGGAGATTTAGAGAGTCTGAGAGATTTCT 746
598 CGGCAAGCTCGACCCCGTCTATCGCGCGGACGAGGAGATCCGCGCGCTGTCGCTCT 657
747 TTCGAGAGAACGAAGAACATCTCTGTCTTATGAGAGCAGGAGTTGTTGAAACAGC 806
658 CTGCGCGCGCACCAAGAACAAACCCCGCTCTATCGCGGAGCGCGCTCGGCAAGACGCG 717
807 TGTGGTTGAGGTTTATGACAAAGAGATTTGAAAGAGAGATGCCCCAAGCTTACTGA 866
718 CGTCTGAGAGGCGCTCGCGAGCGCTGTGCGCGGGAGCTGCGCGCAACCTGCTCGA 777
867 TGTGAGATTAATTTCTGTGACATAGGCTGCTGTAGTTGCTGTGCTTAATACCGAGGAGA 926
778 CGTCCGCTCTGCTCGCTCGACATGAGCGCGCTCTGCTCGCGCGCGCAAGTACCGGCGGA 837
927 GTTTGAAGAAAGGTTGAAATCTCTTTTGAAGAGAGTTGAGGAGCGCTGAAGGCAAGTGTAT 986

838 GTTTCGAGAGCGCTCAAGGCGCTGCTCAAGAGAGTGGAGAGCGCGCAAGGTCAT 897
987 TCTCTTTATGATGATGATTTCTTGTCTGCTGCAAACTGAAGGCTCATGGA 1046
898 CTTCTTCATCGACGAGATACACTCTGCTCGCGCCGAGGAGCGAGGCTCATGGA 957
1047 TGCAGCTAATCTGTTCAAGCCCATGCTAGCTAGAGGCGAGCTTCGATGCTGCTAC 1106
958 CGCGGCCAACCTCTTCAAGCCCATGCTCGCCAGGCGCGAGCTCCGCTGCTCGCGCCAC 1017
1107 AAGCTTTGAAGATACAGGAATATGTTGAGAAAGATGCTGCTTGGAGAGGCTTCCA 1166
1018 CACGCTGGAGAGTACCGCAAGTACGTCGAGAAAGACCGCGCTTCGAGCGGACGCTTCCA 1077
1167 ACAAGTCTATGTTGGGAGCCCAAGTGTGCTGACACCAATTAGTATCTTTCAGAGGACTCAA 1226
1078 GCAAGTGTAGTGGCGGAGCGCGCTGCGCCGACCACTCAGCATCTCTCGAGGCTCAA 1137
1227 GGAGAAATGATGAGGAGCATCATGCTGTGCGAATCCAAGACAGAGCTCTTATTAATGCTGC 1286
1138 GGAGAAATGATGAGGAGCATCATGCTGTGCGAATCCAAGACAGAGCTCTTATTAATGCTGC 1197
1287 TCAAGTGTCTGCTGCTTACATTAATCTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGCT 1346
1198 GCAAGTCTCTGCGCGAGGTATCATCATGCGCGCGCACCTGCGGACCAAGCCCATCGACTGCT 1257
1347 TGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGTATAGTCAACCTTGAAGAGATTTGATTA 1406
1258 GGACGAGGCTGCGCCCAAGCTGAGGCTGAGCTGAGACAGCAGCGCGAGGAGATCGACAA 1317
1407 CTTGAAAGAGAGAGAGATGAGCTGAGAAATTTGAATTTGACGCTTGGAAAGGAGAGAA 1466
1318 TCTGAGCGGAGCGGATTCAGCTGAGGCTGAGCTGCAACGCGCTGGAGAGGAGAGAA 1377
1467 TAAAGCCGCAAGAGCTGAGCTTATAGAGTGGGAGAGAGCTTTGATGACTCGAGAGACAA 1526
1378 CAAGGCGCAGCAGCGCGCGCTGCTGAGGAGTGAAGAGAGCTGAGACACTGAGGAGCAA 1437
1527 GCTTCAAGCTCTCAAGTGAATGAGAGAGAGAGAGAGATTTGATGATTTGATTTGAG 1586
1438 GCTGCGCGCTGCTGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGATCGAGATCCGGA 1497
1587 GCTTAAACAGAAAGAGAGAGCTGAGTCAATGTTTCTTGGAGGAGAGAGAGAGAGATATGA 1646
1498 GCTGAGAGAGCGCGGAGAGAGCTGAGTTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1557
1647 CTTGCAAGAGCTGCTGATCTAAGATATGCGCAATTTCAAGAGAGTGAATCTGCAATTCG 1706
1558 CTTGCGCACGCTGCGCGACCTCAAGTACGCGCGCTTCAGAGAGATCGACGCGCCATTCG 1617
1707 CCAACTTTGAAGAGAGAGCTTCTTGAAGAGAGATGATGCTCAGAGAAAGAGAGAGAGAG 1766
1618 GAAGCTGAGAGG---CGAGACCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
1767 ACATGCTGCTGAGGTTGAGCGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1826
1675 GCAGATCGCGGAGGTTGAGCGCTGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
1827 TGAGAGAGAGAGAGGTTGATGCTGCTGATGAGGTTGATGAGAGAGAGAGAGAGAGAGAGAG 1886
1735 GCAGAGAGAGAGAGGTTGCTGCGCATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
1887 TCAAGCGGTAATGAGGTTTCTGAGGCAATTTCAAGGTCAGGAGAGAGAGAGAGAGAGAGAG 1946
1795 GGAGCGCTGAGCGCTGCGCGAG 1854
1947 ACAAGCGCAACTGAGTATCTTATCTTGAACCACTGCTGTTGAGGAGAGAGAGAGAGAG 2006
1855 GCAG 1914
2007 CGCCAG 2066

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato; *Arabidopsis thaliana*; ds.

1498	TGACGCAGCTGTGCAAGATGCTGATCCAGGAGGAGATCGACGAACTGCAACGCTCTACA	1557
2705	TAGATGCAGCGCGTGGT-----GATCTTGTGTACCGGTAGA---AAGTGGAGGTCTAG	2755
1558	TCGACGCCGCGCCGGTAAAGGACGAACCTGGTCTACAGGCTGACCGGAGCGCGTCTGG	1617
2756	TGGACGCTTCAACAGGCAAGAGTCAAGTGTCTGATTCATATTGCTTAACGGGCCAAGA	2815
1618	TGAACGCTGAGACGGGGATGAAGTCGGAATCCTGATCAAGTCCCAACCAAGCTCCACCA	1677
2816	GAACTGATGCAGCTTCAGGCGGTCAAGAAATCAGGATCGAGAAATACAGATGACGATA	2875
1678	GGACGACGCTGCGCAGCGCGTCAAGAAATCAGGATCATGGAGGAGGACGAGGACGCA	1737
2876	ATGAGGA 2882	
1738	TGGACGA 1744	
RESULT 11		
AAI66072		
ID	AAI66072 standard; DNA; 2045 BP.	
XX	AAI66072;	
XX	14-JAN-2002 (first entry)	
DE	Zea mays heat shock protein 101 gene (GenBank: AF077337).	
XX	Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;	
KW	canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;	
KW	Arabidopsis thaliana; ds.	
OS	Zea mays.	
XX	WO200170929-A2.	
XX	27-SEP-2001.	
XX	20-MAR-2001; 2001WO-US08836.	
XX	20-MAR-2000; 2000US-190769P.	
PR	18-APR-2000; 2000US-198116P.	
XX	(ARCH-) ARCH DEV CORP.	
PA	Lindquist S, Queitsch C, Vierling E;	
PI	WPI; 2001-639123/73.	
DR	P-PSDB; AAM51663.	
XX	Transgenic plants with improved heat stress tolerance, useful for	
PT	producing animal feed, oil and synthetic products -	
XX	Claim 4; Page -; 91pp; English.	
PS	The invention relates to a transgenic plant, comprising a genetic	
XX	construct comprising a promoter operatively linked to a nucleic acid	
CC	sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)	
CC	family amino acid sequence (AAM51651-AAM51671). The transgenic plant has	
CC	increased stress tolerance, especially to heat. The plant is a cereal,	
CC	grass, ornamental plant, crop plant, food plant, oil-producing plant, a	
CC	synthetic product-producing plant, an environmental waste absorbing	
CC	plant, an alcohol plant, a medicinal plant, a recreational plant and/or	
CC	an animal feed plant. In particular, the transgenic plant is cotton,	
CC	canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or	
CC	Arabidopsis thaliana. The plants may be used to produce animal feed,	
CC	alcohol, crop, oil, medicine or a synthetic product.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained from GenBank using the Accession Number	
CC	reference provided in the specification.	
XX		

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51671.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA16057-AA16084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3942 BP; 834 A; 1111 C; 1236 G; 760 T; 1 other;

Query Match 26.0%; Score 807; DB 22; Length 3942;

Best Local Similarity 58.5%; Pred. No. 6.6e-212;

Matches 1918; Conservative 0; Mismatches 845; Indels 514; Gaps 9;

117 CGAGTCTCATATATCGTTTAAAGGATTAACAAAGCTAATCGAAGATGAATCCAGAA 176
390 CCGCTTGTGTAGTAGTTGTAGGACTGAGGACCGAGAGCAGCCATGAATCCGACAA 449
177 ATTACACACAAACAGACAAATGCTACAGCTCATGACTAGCTGTGAATCCAGG 236
450 CTTACACCAACAGACGACGAGGATGCTGGGGGCGCAGATGTCGGTGGAGCCGG 509
237 ACATGCTCAATTCACCTCTTTGCAATTTAGCTGTGTGCTTGTATCTGTATCCCA 296
510 CCAGCGCAGCTCAAGCGGCTGACCTGGCGCAGTGTGCTGGCGACAAAGGGCGGCAT 569
297 ATTTCTCAAGCATCTTAGTCCGCTGGCGGAGACGACGAGC-----TCAATCTCTGA 350
570 CTTGGCGAGCCCATCAGGGGGCGTCCGGGGGCGACGAGCGCGGGGACTCTGTCCA 629
351 AAGAGTGATCAATCAAGCCCTTGAAGAGCTTCTCTTCAATCTCTCTCCACTGATAT 410
630 GCGGCTGCTGAACACTCGCTCAAGAGCTGCGCTGCGAGTCCCGCGCGGACTCCGT 689
411 TCCAGCGAGTTCTAGTCTTATTAAGTCAATCTGCTGCTGCTCAAGCTGCTCAGAGTCA 470
690 TCCGCGCTACCGCGCTGATCAAGGTCAATCCGCGGCGCAGTCCGCGCAGAGAAACG 749
471 AGGTGATCTATTTGCTGTGACCAAGTTGATTTATGSGTCTTCTTGAAGATCTCAAT 530
750 CCGGACTCTCGACCTCGCGCTGACACCACTGCTGCTCGGCTGCTCGAGACTCGGAT 809
531 CAGGATTTGTTGAACAGAGTCTGTAGCGACCGCGGAGGCTAAAGTCTTGAGGTTGAA 590
810 CTCGACTGCTCAAGAGGCGCGGCTGCTCGCGGCGGCTGCGCGCGGCTTGAA 869
591 GCTTCTGGGAAAGAGGAAAGTTGAGAGTGTCTTTCAGGGGACACAAATTTTCAAGC 650

870 GCTCCGCGCGCGGAGCGCGCGCTGGAGTCCGCGTCCGGGGATACCAACTTCCAGGC 929
651 TTTAAGACTTATGGAAGAGATTGTTGAGCAAGCAGGAGGAGCTTGAATCTCTGTGATGG 710
930 GCTCAAGACATACGCGCGGACCTCTGTCGAGCAGCGCGGAGAGCTTGAATCCGCTCATCG 989
711 TCGTGATGAGGAGATTAGAGAGTCTGTCGAGGATTTCTTTCAGGAGAACGAAACAATCC 770
990 CCGGACGAGGAGATCCGCGCGCTGTCGCGCAATCTCTCGCGCGCACCAAGAATAACCC 1049
771 TGTGCTTATTTGAGAGCCAGGAGTGTGTAAGACAGCTGTGTTGAAGTTTACACAAAG 830
1050 CGTCTCTCATCGCGGAGCCCGCGCTTGGCAAGACGCGCGCTGTCGAGGCGCTCCGCGAGC 1109
831 GATTGTGAAGAGAGATGTCGCCAACAGTCTTACTGATGTGAGATTAAATTTCTGTTGACAT 890
1110 CATGCTTCGCGGAGATGTCGCCAGTAACTCTCTCGAGCTCCGCTCATCGCGCTCGACAT 1169
891 GGGTCGCTTAGTTCTGCTGCTTAAATACCGAGGAGATTGGAAGAGGTTGAAATCTGT 950
1170 GGGCGCTGCTGTCGCGGCGCAAGTACCGCGCGAGTTCGAGGAGCGGCTCAAGCCGT 1229
951 TTTCAAGAGATTGAGGAGCTGAGCGCAAGTATCTCTTTATTGATGAGATTCAATTT 1010
1230 GCTCAAGGAGTGAAGAGCGCGGAGGAGGATCTCTCTTCTCATCGACGAGATACACT 1289
1011 GGTTCCTGCTGCTGCGCAAACTGAAGGCTGATGATGATGATGATGATGATGATGATGAT 1070
1290 GCTCTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1349
1071 GTTAGCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1130
1350 GCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1409
1131 TGTGAGAGAGATGCTGCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1190
1410 GGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1469
1191 TGTGCTGCTGACACCAATTTAGTATCTTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1250
1470 GGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1529
1251 TGTGAGGATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1310
1530 GGTGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1589
1311 TGTGCGGATTTACCGGATTAAGCAATTTGATTTGATGAGGAGGAGGAGGAGGAGGAGGAG 1370
1590 GGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1649
1371 AGTCCAGCTTGAATGATCACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1430
1650 GGTGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1709
1431 GGAATTTCAATCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1490
1710 TGTGAGGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1769
1491 AGAGGTGC----- 1498
1770 TGTGAGGATGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1829
1499 -----GGAAGAG 1506
1830 GTTTGAACTTTTGATTAACGTCGCTTAACCTGTCGCGCATTTGTCAGGTCAGGAGGAA 1889
1507 CTTGATGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1566
1890 TTTGAGGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1949
1567 AGAATTTGATGAGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1626

1950 AGAATTGATGAGATCAGGAAGCTGAAGCAGCGCGGAGAGCTCCAGTTTACCCCTGCAG 2009
1527 GAGCAGAAAGATATGACCTTGCAGAGCTGCTGATCTTAAGATATGGCGCAATTCAA 1686
2010 GAGGCCGAGGCCCGGATGAGCTGCGCCGCTGTGCGCCGACCTCAAGTAGCGGCCCTCCAG 2069
1687 GAAGTGAATCTGCAATGCCCAATCTGAAGAACTTCTTCTGAAGAGAAATGTGATGCTC 1746
2070 GAATTCGACGCTGCTATCTCAAGTGA---GAGCGAAACAGGGGAGAACCTGATGCTC 2126
1747 ACAGAAACGTTGGCCCTGAACACATTTGCTG----- 1777
2127 ACCGAAACCGTGGCCCTGAACAAATTCAGAGGTATGTTATTCTTGTTCACCGTCA 2186
1778 ----- 1777
2187 CAATAATTTGCAGAGCAAGTGCAGAAATTCGCCGATCGTCTCCTAGTAGAGTAGTCTGTC 2246
1778 -----AGTTGTGAGCCGTTGGACAGGATTC 1805
2247 AGCGTGCTGAATGGTGTTCGTCCTATGCGCAGGTGTGAGCCGTTGGACGGGTATTCC 2306
1806 AGTGACGAGACTTGGCCAAATGAGAGAGAGAGGTGATGCTCTTGTCTGATAGGTTGCA 1865
2307 AGTGACCGGCTTGGCCAAACGACAGAGAGAGGCTGGTTGGCCCTGAGTGGCTTCA 2366
1866 TAAGCGGCTTGTGGACAGAACTCAAGCGTAAATGTCAGTTTCTGAGGCAATTCCTAAGTTC 1925
2367 CCAGAGGCTGTGGCCAGACAGAGGCTGTGAGCGCCGTGCGAGAGCGGTGTCTGAGGTC 2426
1926 AAGGCAGAGCTTGTGAGGCAACAGCCCACTGGATCATTTCTTATTCCTTGGACCAAC 1985
2427 GAGGCCGCTTGTGGCAGCCACAAACAGCCACTGGCTCGTTCTCTCTGGGTCGAC 2486
1986 TGGTGTGGCAAACTGAGCTCGCAAGCTCTGCTGAGCAGCTGTTGATGATGAAGA 2045
2487 TGGCTGGGAAACTGAGCTGGCCAGGCGCTTAGCCGACAGCTGTTTCGACGAGAGAA 2546
2046 CTTCTTAGTTCGATTTGATATGTCGGAATATATGAACAACTCTGCTCTCGCCCTCAT 2105
2547 CTTCTTGTCCGATTCGACATGTCGGAGTACATGGAGCAGCACTCGGTTGCCGCTCAT 2606
2106 TGGGCCACCAACAGGGTATGTT----- 2127
2607 CGGACCAACCTGGTGAAGTAGCAGAAAATGATGATCATCTTGGTATTTTAACTGGAA 2666
2128 -----GTCACAG 2136
2667 TTGCGATGAACCTTTGTTCTGACAGTGGGCCGACCTGTGACAGCTACGTCGCCATGAA 2726
2137 GAAGGTGACAACTAACTGAGGCTGTGAGGAGGACCTTATTGTGTCATACTCTTTGAT 2196
2727 GAGGTGGCAGCTGACTGACAACTGAGGAGGAGCGGTACAGCGTATCTGTTGAC 2786
2197 GAAGTGAAGAGGCTCATGTTGCTCTTCAACTCTGCTCCAAAGTTTTCGATGATGGT 2256
2787 GAGTTCGAGAGGCGCATGTGCGCGTGTTCACACCTGCTCCAGGTCCTCGACGACGCG 2846
2257 CGATTGACAGCGGCGAGGAGGAGCAGTGCATTTTCAGAACTCGGTGATATCATGACA 2316
2847 AGGCTGACGGATGGGCAAGGCGAGGACGCGTGGACTTCAGGAAACCGGTGATCATGACA 2906
2317 TCAGAACTTGGTGTGACACCTCTCTGCGAGGCTAACTGGGAAAGTAACTGGAAGTGTG 2376
2907 TCGAACCTCGGCGCGAGCACCTCTCTGCTGGGATGGTGGGCAAGAACTCCATGAGGTC 2966
2377 GCCCGGACTGTGTGATGCGG----- 2398
2967 GCTCGCATCTGCTCATGACGAGGTATGCAATTGCGATAACGTTCTGATCGTCTGATAGA 3026
2399 -----AGGTGAGGAA 2408
3027 CATTTGCTCTGCAGCGTATGCTCAATCGGATCGTTATTGCTGTGTGACGGTGGAGG 3086

QY 2409 ACACCTTCAGACAGAGCTCTTGAACAGGCTTTCACAGAGATTGTGGTTTCGACCCCTTTC 2468
Db 3087 GCACTTCCGCCCTGAGCTGCTGAACCGTCTCCAGAGATCGTGAATCTTCGATCTCTGTC 3146
QY 2469 ACATGACAGTTGAGAGAAAGTAGCTCGGCTTCAATGAAGAGCTGTCTGTCGGGTTGC 2528
Db 3147 CCACGACAGCTGAGAGAGGCTCGCTCGCCTTCAGATGAAGGATGTGGCGCTCTGTTGC 3206
QY 2529 TGAAGAGAGAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTGCAGAGAG 2588
Db 3207 CGAAGGGGATGCGCTCTGCTGTGACCGACCGCATTTGGAATCATCTTGTCTCTCTC 3266
QY 2589 TTATGACCC----- 2597
Db 3267 TTACGATCCGGTATGTGACCATCCATGATTTGATCCATCTGAATTCGTGCGTGAACCTTG 3326
QY 2598 -----GCTGATGCTGTAGC 2614
Db 3327 ATGGTCTGACTCTCTTATCTTTCTTTGTGTGCTTCAACACAGGTTGATGCGCGCGC 3386
QY 2615 CTATAGGAGATGAGTGGAGAGAAAGTGTGTCACAGAACTGTCAAGATGTGTGCGGTG 2674
Db 3387 CAATCAGGAGGTGATCGAGAGAGAGGTTGTGACGAGCTGTGGAAGATGCTGATCCAG 3446
QY 2675 AGAAATTCGATGAAATCTCCACTGTTTACATAGATG-----CAGCGCTGTGTGATCTTG 2728
Db 3447 AGGAGATCCAGAGAACTGACCGTCTACATCGACGCGCGCCGCAAGGACGAGCTGG 3506
QY 2729 TGTACCGGTTAGA---AGTGGAGGCTAGTGGACGCTTCAACAGGCAAGATGATGATG 2785
Db 3507 TCTACAGGTTGGACCGGAGCGCGGCTCTGTTGAACGCTGAGACGGGATGAAGTCGACA 3566
QY 2786 TGCTGATTCATATTGCTTAACGGGCCAAAGAGAGATGATGATGATGATGATGATGATGATG 2845
Db 3567 TCTGATTCAGGTTCCCAAGCTCCACAGGAGGACGCTGCGGAGGCGGCTGCAAGAGA 3626
QY 2846 TGAGGATCCAGGAAATAGAGATGACGATTAATGAGGA 2882
Db 3627 TGAGGATCATGAGGAGGACGAGGCGGATGAGGA 3663

RESULT 13

AAI66074

ID AAI66074 standard; DNA; 3942 BP.

XX AC AAI66074;

XX DT 14-JAN-2002 (first entry)

XX DE Triticum aestivum heat shock protein 101 gene (GenBank: AF097363).

XX KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

XX KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

XX KW Arabidopsis thaliana; ds.

XX OS Triticum aestivum.

XX PN W0200170929-A2.

XX PD 27-SEP-2001.

XX PF 20-MAR-2001; 2001WO-US08836.

XX PR 20-MAR-2000; 2000US-190769P.

XX PR 18-APR-2000; 2000US-198116P.

XX PA (ARCH-) ARCH DEV CORP.

XX PI Lindquist S, Queitsch C, Vierling E;

XX XX WPI; 2001-639123/73.

XX DR P-PSDB; AAM51660.

117 CGAGTTCATATATATCGTTTAAAGGATTTACAAAGCTAATCGAAGATGAATCCAGAGAA 176
330 CCGCTGTGTTAGTAGATGTTGTGAGACTTGAGGACCGAGAAGACCATGAATCCGGCAA 449
177 ATTACACACAGACAAACGAGACAAATTGCTACAGCTCATGAGCTAGCTGTGAATGCAGG 236
450 CTTTACCCACAGACGAACGAGGCGCATCGTGGGGCGCACAGAAATTGCGGTGAGGCGCG 509
237 ACATGCTCAAATTCACCTCTCTTGCATTTTAGCTGGTCTTTGATCTCTGATCCCAACCGTAT 296
510 CCACGCGAGCTCACCGCGCTGCACCTGGCCGAGTGTGCTGCGGACGAAGGGCGCAT 569
297 ATTTTCTCAAGCAATCTTAGTGC CGGTGGCGAGAACGACG-----TCAATCTGCTGA 350
570 CTTGGCGGACGGCCATCAGGGGGCGTCCGGGGGGGACAGGAGCGCGCGGGAGCTCTGTTGGA 629
351 AAGAGTGATCAATCAAGCCTTGAAGAAGCTTCCTTTCACAATCTCTCCACCTGATGATAT 410
630 GCGCGTGTGAACAATCGCTCAAGAAGCTGCGGTGCGAGTCCCGCGCGCGGACTCCGT 689
411 TCAGCGAGTTCATGTTTAAAGTCAATTCGTGCTCTCAAGCTGCTCAGAAGTCAAG 470
690 TCCGCGCTCAGCGCGCTGATCAAGGTATCCGCGGGCGAGTCCCGCGCAGAGAAACG 749
471 AGGTGATACCTATTTGGCTGTGACAGTTGATTATGGGTCTTTGAAGATTTCTCAAT 530
750 CGGGGACTCGACCTCGCGTGCACAGCTGCTCTCGGCTGCTCGAGGACTCGCAT 809
531 CAGGATTTGTTGAACAGAGTCCGTGTAGCGACGCGGAGGTAAAGTCTGAGGTTGAGAA 590
810 CTCGACTCCCTCAAGAGGCGCGGCTGTCCCGCGCGGGGTGCGCGCGAGCTTGAGAA 869
591 GCTTCGTGGGAAGAGGGAAGAAAGTTTGAGAGTCTTCAGGGGACACAAATTTTCAAGC 650
870 GCTCCGCGGGGAGGGCGCGCGTGGAGTCCGCTCGGGGATACCAATCTTCAGGC 929
651 TTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAGCTTGATCTCTGTGATGG 710
930 GCTCAAGACATACCGCGGGACCTCTCGTCAAGCAGGCGGGGAAGCTTGACCCCGTCAATCG 989
711 TCGTGATGAGGAGATTAGAAGAGTCTGAGGATTTCTTCGAGAGAACGAGAGCAATCC 770
990 CCGCGCAGGAGATCCGCGCGTCTGCGGCAATCTCTCGCGCGGCAACAAGATAACCC 1049
771 TGTGCTTATGAGAGACCGAGGAGTTGGTAAAAACAGCTGTGGTTGAAGCTTTAGCACAAAG 830

2127 ACCGAAACCGTCGCGCCCTGAACAAATTGCAGAGTATGTTATTATTCTTGTTCACCGTCA 2186
1778 ----- 1777
2187 CAAAAATTTTCAGAGCAAGTGCAGAAATTTGCCGATCGTCTCCTAGTAGAGTAGTCGTGC 2246
1778 -----AGTTGTGAGCCGTTTGACAGGATTC 1805
2247 AGCGTCTGAAATGGTGTTCCTTATGGCAGGTGCTGAGCGGTTGACGGGTATTC 2306
1806 AGTGACGAGACTTGGCCAAATAGAGAGGAGGTTGATGTTGCTTCTGTGATAGGTGCA 1865
2307 AGTGACCGGCTTGGCCAGAACGACAGAGAGGCTGTTGGCTCTGCTGACAGGCTTCA 2366
1866 TAGCGGTTTGGGACAGATCAAGCGTAAATGCAAGTTCTGAGCRAATCTTAAGTTC 1925
2367 CCAGAGGTTGTCGCGCAGACAGAGGCTGTGAGCGCTGCGAGAGCGGTGCTGAGGTC 2426
1926 AAGGCGAGGACTTGGTAGGGCACAAACAGCCAACTGGATCATCTTATTCCTTGGACCAAC 1985
2427 GAGGCGGCTTGGCAGGCCACAAACAGCCCACTGGCTCTCTCTCTGCTGGGTCCGAC 2486
1986 TGGTGTGGCAAACTAGCTCGCAGAGCTTCTGTCAGCAGCTGTTTGCATGATAA 2045
2487 TGGCGTGGGAAACTAGCTGGCCAAAGGCGCTAGCCGAAACAGCTGTTTCAGACGAGAA 2546
2046 CCTCTAGTTCCGATTGATGTCGGAATATATGGAACAACTCTGCTCTCGCTCAT 2105
2547 CCTCTTGTCCGATCGACATGTCGGAGTACATGGACGACACTCGGTTCCCGCTCAT 2606
2106 TGGGCGCCACACAGGGTATGTT----- 2127
2607 CGGAGCACCACTGGTAAAGTAGCAGAAATGCATGCTCTGTTATTTTAACTCGAA 2666
2128 -----GGTCAGAG 2136
2667 TTGCGATGAATCTTGTCTGACAGTGGCGGACCTGTGACAGCTACGTCGCCATGAA 2726
2137 GAAGTGAGCAACTAACTGAGGCTGTGAGGAGCGGCTTATTGTGTGCTACTCTTTGAT 2196
2727 GAGGTGGGAGCTGATGTAACAGTAGAGGAGGCGGTACAGCGTATCTGTTCCGAC 2786
2197 GAAGTGAGAGGCTCATGTTGCTGCTTCAACACTCTGCTCCTCAAGTTTGGATGATGTT 2256
2787 GAGTTCGAGAGGCCCATGTCGCGCTGTTCACACCTGCTCCAGGTCCTCGACGCGC 2846
2257 CGATTGACAGCGGCAAGGACGACAGTGAATTTGAGAACTCGGTGATTAATCATGACA 2316
2847 AGGCTGACGGATGGGCAAGGACGAGCGTGACTTCAGGAACACCGTGATCATCATGACA 2906
2317 TCAAACTTGTGCTGAACACTCTTTCAGGCGCTAACTGGGAAAGTAAACATGGAAGTG 2376
2907 TCGAACCTCGCGCGGAGCACTCTCTGCTGGGATGTTGGGCAAGAACTCCATGAAGTTC 2966
2377 GCCCGGAGCTGTGTGATCGGG----- 2398
2967 GCTCGGATCTGGTCAATGACGAGGATGATGATTCGATTAAGCTTCAATCTGTTAGA 3026
2399 -----AGGTGAGAA 2408
3027 CATTTGTTCTCTGAGCGGTGATGCTCAATCGGATCGTTATTGCTGTGTGAGGTTGAGG 3086
2409 ACATTTACAGACAGAGCTTTGAACAGGCTTGACGAGATGTTGTTGTTGACCCCTTTTC 2468
3087 GCATTTCCGCTGAGCTGTGAACCGCTCTGACGAGATCGTGAATCTGATCTCTGTC 3146
2469 ACATGACAGTTGAGGAAGTAGTTCGGCTTCAATGAAGAACGTTGCTGTCGGCTTGC 2528
3147 CCAGGACAGCTGAGGAAGTTCGCTTCCAGTGAAGATGTTGGCGGCTCTCTTTC 3206
2529 TGAAGAGAGGTTGTTGGGAGTCACTGATGCTGCTTGGACATATATCTTGGCAGAGAG 2588
3207 CGAAAGGGGCACTCGCTCTGGTGTGACCGGCGCATTTGGACATCATCTTGTCTCTC 3266

QY 2589 TTATGACCC----- 2597
DB 3267 TTACGATCGGTATGTGACCATCCATGATTTGATCCATCTGAATTCGTCGTCGACACCTG 3326
QY 2598 -----GGTGTATGTCGTAGC 2614
DB 3327 ATGGTCTGACTCTCTTATCTTCTTGTGTGCTTCAACACAGGTGTATGCGCGCGC 3386
QY 2615 CTATAGGAGATGAGTGAAGAGAGGTGTCACAGACTGTCAAGATGTTGTGCGTG 2674
DB 3387 CAATCAGAGGTGATCAGAGAGGTTGGTACGAGCTGTGGAAGATGCTGATCCAGG 3446
QY 2675 AGGAAATCGATGAAATCCACTGTTTACATAGATG-----CAGGCGCTGCTGATCTTG 2728
DB 3447 AGGAGATCGACGAGAACTGACGGTCTACATCGACGCGCGCGCAAGGACGAGCTGG 3506
QY 2729 TGTACCGGTAGA-----AAGTGGAGTCTTAGTGAGGCTTCAACAGGCAAGAGTCAGATG 2785
DB 3507 TCTACAGGTGACCGGAGCGCGCTCTGTTGAACGCTGAGACGCGGATGAAGTCGACA 3566
QY 2786 TGCTGATTCATATTCTTAACGGGCCAAAGAGAGATGTCAGCTCAGGCGCTGAAGAAGA 2845
DB 3567 TCTGATCCAGTCCCAACAGCTCCACAGAGGAGCGCTGCGCAGGCCCTCAAGAAGA 3626
QY 2846 TGAGGATCGAGAAATAGAAGATGACGATAATGAGGA 2882
DB 3627 TGAGGATCATGAGGAGGACGAGGACGCGCATGGACGA 3663

RESULT 14

AA13093
ID AA13093 standard; DNA; 6491 BP.
XX AC AA13093;
XX DT 19-MAR-1999 (first entry)
XX Enterococcus faecalis genome contig SEQ ID NO:156.
DE Enterococcus faecalis; contig; detection; Enterococcal infection;
KW Enterococcus faecalis; contig; detection; computer readable medium; ds.
XX vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Dillion PJ, Kunsch CA;
XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX Claim 1; Page 885-888; 2084pp; English.
XX A computer readable medium has been developed which has recorded on it
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AA12938 to AA13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with

commercial importance. The products can be used to detect the presence of *Enterococcus faecalis* in samples. They can also be used for diagnosing *Enterococcal* infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, in vivo or in vitro. In particular the polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection.

Sequence 6491 BP; 2164 A; 1114 C; 1351 G; 1844 T; 18 other;

Query Match 21.7%; Score 672.8; DB 20; Length 6491;
 Best Local Similarity 59.1%; Pred. No. 1e-174;
 Matches 1180; Conservative 5; Mismatches 807; Indels 6; Gaps 2;

683 AAGCAGGAGGCTTCAATCTGATGTTGGTGTGATGAGGAGATTAGAAAGATCGTAGGA 742
 |||||
 546 AAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 605
 |||||
 743 TTCTTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 802
 |||||
 606 TTTTATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
 |||||
 803 CAGCTGTGCTTGAAGGTTTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 862
 |||||
 666 CAGCGATTGTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 725
 |||||
 863 CTGATGTGAGATTAATTCCTGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 922
 |||||
 726 AAGATAAAACCAATTTTCTTTAGATATGGGCGCTTAATTTCGGGAGGAGGAGGAGGAGG 785
 |||||
 923 GAGAGTTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982
 |||||
 786 GCGAATTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 845
 |||||
 983 TGATTCCTTTATGATGAGATTCATTTGGTTCTTGGTCTGGCAAAACTGAAGGAGGAGG 1042
 |||||
 846 TCATTTTATTCATGATGAGATTCATTAATATCGTCGGAGCTGTTAAACAGAGGAGGAGT 905
 |||||
 1043 TGATGACGATTAATCTGTTCAAGCCCATGTTAGCTAGAGGAGGAGGAGGAGGAGGAGGAG 1102
 |||||
 906 TGACGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 965
 |||||
 1103 CTACACGCTTGAAGATACAGGAGATATGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1162
 |||||
 966 CCACACGCTTGATGAGATATCGCAATATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1025
 |||||
 1163 TCACACAGGCTATGTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1222
 |||||
 1026 TCCAAAAGGATTTAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1085
 |||||
 1223 TCAGAGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1282
 |||||
 1086 TAAAGAGAGGATTTGAATTTACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1145
 |||||
 1283 CTGCTCAGCTGTCTGCTGTTATCAATACCTGTCGGCATTTACCGGAGGAGGAGGAGGAGG 1342
 |||||
 1146 CCCTACTTTATCCGATCGCTATATTTACTGATGCTTTCTTACAGATTAAGCGATGACT 1205
 |||||
 1343 TGGTTGATGAGGCTTGTGGGAATGTGAGAGTCCAGCTTGTATGATCAACCTGAGAGATTTG 1402
 |||||
 1206 TAGTCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1265
 |||||
 1403 ATACCTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1462
 |||||
 1266 ATCAGTAAACAGCTCGGTTAATGMACTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1325
 |||||
 1463 AGGATAAGCCAGCAAGCTCGACTTATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1522
 |||||
 1326 CAGATGATGAGTAAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1385
 |||||
 1523 ACAAGCTTCAGCCTCTCACGATGAATATACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1582

1386 AAGAGCCCACTCAATGAAATGCAATGGAAACGAAACGAAAGAGTTAATGCGGTTT 1445
 |||||
 1583 GAGGCTTAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1642
 |||||
 1446 CCAATAAGCGTGAGAGATGTAAGAGCAAAACAGAAATAGAGATGCTGAAATATTT 1505
 |||||
 1643 ATGACCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1702
 |||||
 1506 ATGATTTAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1565
 |||||
 1703 TTCCCAACTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1762
 |||||
 1566 TGAAGAGATTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1625
 |||||
 1763 CTCAACAC---ATTGCTGAGGTTTGTGAGCGGTTGAGACGGGATTCAGTGACGAGAGTTG 1819
 |||||
 1626 CTGAAACGAAATTTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1685
 |||||
 1820 GCCAAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1879
 |||||
 1686 TTGAGGCGGAAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1745
 |||||
 1880 GACAGATCAGCGGTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1939
 |||||
 1746 GTCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1805
 |||||
 1940 GTAGGCGCACAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1999
 |||||
 1806 AAGATCCAAATCGCCCACTCGGTTCTGTTCTTAGGACCAACTGTTGGTAAATA 1865
 |||||
 2000 CTGAGCTCGCAAGGCTTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2059
 |||||
 1866 CAGAACTTGCTAAAGCTTTAGCTGAAAGATTTGTTGATTTCTGAAGATCATATGGTACGGA 1925
 |||||
 2060 TTGATATGTCGGAATATATGGAACACACTCTCTCTCTCGCTCATTTGGGGCACACACAG 2119
 |||||
 1926 TTGACATGATGAGATATCATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1985
 |||||
 2120 GGTATGTTGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2179
 |||||
 1986 GCTATGATGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2045
 |||||
 2180 GTGTCTATCTCTTTGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2239
 |||||
 2046 CAATGTCTTATTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2105
 |||||
 2240 AAGTTTGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2299
 |||||
 2106 AAGTTTGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2165
 |||||
 2300 CGGTGATAATCATGACATCAAACTTGGTCTGAAACACCTCTTGCAGGAGGAGGAGGAGGAG 2359
 |||||
 2166 CAGTTTAAATTTATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2225
 |||||
 2360 AAGTAAACAAAT---GGAGTGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2416
 |||||
 2226 AAGGAAACAAATTTCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2285
 |||||
 2417 GACCAGAGGCTTTGAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2476
 |||||
 2286 AACAGAGGTTCTTTAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2345
 |||||
 2477 AGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2536
 |||||
 2346 ATGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2405
 |||||
 2537 GAGTTGCTTTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2596
 |||||
 2406 AAATGTTGTTGGAATCACTGATGAAGCTAAACATGAGATTTGCTGAAATTTGGGTATGAGC 2465
 |||||
 2597 CGGTGATGCTGCTAGGCTTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2656
 |||||

2466 CAGCTTTAGGAGCTGCTCGTTAAACAGTTTCATTACTCGGAGTGGAAACACCACTTAG 2525

2657 CAAAGATGTTGTGGTG 2674

2526 CAAAAGAAATGTTCTG 2543

LT 15

8988

ABS98888 standard; DNA; 6491 BP.

ABS98888;

18-DEC-2002 (first entry)

Enterococcus faecalis contig sequence #156.

Computer readable medium; Enterococcus faecalis; microbe; growth; pathogenicity; vaccine; resistance; Enterococcal infection; commercial; therapeutic; industrial; fermenting; sugar source; metabolite; vaccine; biotech technology; antibacterial; modulator of nucleic acid expression; contig; ds.

Enterococcus faecalis.

US2002120116-A1.

29-AUG-2002.

04-MAY-1998; 98US-0070927.

04-MAY-1998; 98US-0070927.

(KUNS/) KUNSCH C A.

(DILL/) DILLON P J.

(BARA/) BARASH S.

Kunsch CA, Dillon PJ, Barash S;

WPI; 2002-750065/81.

Computer readable medium having recorded on it a Enterococcus faecalis nucleotide sequence useful for detecting diseases related to Enterococcus infections in animals

Claim 1; Page -; 119pp; English.

The present invention relates to a new computer readable medium with an Enterococcus faecalis nucleotide sequence. The invention is useful to diagnose the presence of E.faecalis in a sample or determining the presence of a specific microbe in a sample. The invention is also useful for modulating the growth or pathogenicity of E.faecalis, in a vaccine to confer resistance to Enterococcal infection, for commercial, therapeutic and industrial purposes, and for fermenting a particular sugar source or to produce a particular metabolite. The invention is useful for detecting diseases related to Enterococcus infections in animals, and for detecting E.faecalis using biotech technology. The present nucleic acid sequence represents an Enterococcus faecalis contig DNA sequence of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at <http://seqdata.uspto.gov>.

Sequence 6491 BP; 2164 A; 1114 C; 1351 G; 1844 T; 18 other;

Very Match 21.78; Score 672.8; DB 24; Length 6491;

Local Similarity 59.18; Pred. No. 1e-174; Matches 1180; Conservative 5; Mismatches 807; Indels 6; Gaps 2;

683 AAGCAGGAGGAGCTTGATTCCTGTTGATTCGTTGATGAGGAGATTAGAGAGTCTGTGAGGA 742

546 AAGCAGGAGGAGGAGGAGGTTGATTCGTTGATGAGGAGGTTGATTCGTTGATGAGGAGGTTGATTCGGA 605

QY 743 TTCTTTTGAGGAGAACGAGAACAACTCTGTGCTATTATGGAGAGCCAGAGTGTGGTAAAA 802
 DB 606 TTTTATCAAGAAAACTAAAAATAATCCAGTCTTAATTTGGTGAACCAAGTGTGGTAAAA 665
 QY 803 CAGCTGTGTTGAAGGTTTAGCCAAAGGATTGTGAAGGAGATGTGCCCAACAGTCTTA 862
 DB 666 CAGCGATTGTTGAAGGATTAGCACACGAATCGTTCGTAAGATGTTCCCGAAACTTAA 725
 QY 863 CTGATGTGAGATTAAATTTCTGTTGACATGGTGGTGTGTTAGTTGCTGTGCTAAATACCGAG 922
 DB 726 AAGATAAAACCAATTTTCTTTAGATATGGCGCTTAATTTGCGGAGCAAAATTCCTGTG 785
 QY 923 GAGAGTTGAAGAAAGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGAAGGCAAG 982
 DB 786 GCGAATTTGAAGAACGGTTAAAGCTGTTTTAAAGAAAGTGAAGAAAGTGTGCGCAAAA 845
 QY 983 TGATTCCTTTTATTGATGAGATTCAATTTGTTGTTGCTGTGCGCAAACTGAAGGTCGA 1042
 DB 846 TCAATTTTATTCATTTGATGAATTCATTAATTCGTGCGAGCTGTGTAAACAGAGGCGAGTA 905
 QY 1043 TGGATGAGCTTAATCTGTTCAAGCCCATGTTTAGCTAGAGGCGAGCTTCGATGCAATTCGTTG 1102
 DB 906 TGGACGCGGAAATTTTATTAACCGATGCTAGCACGTGGTGAATTCATTTAATTCGTTG 965
 QY 1103 CTACAAAGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTTGAGAGGAGGT 1162
 DB 966 CCACAAAGCTTTGATGAATATCGCAATATATGGAAGAAAGATAAGAGCTTTAGAACGTCGTT 1025
 QY 1163 TCCAAAGCTCTATGTTGCGGAGCCAAAGTGTGCGCTGACACCATTTAGTATCTTTAGAGGAC 1222
 DB 1026 TCCAAAGCTATTTAGTCAAGAACCAACAGTAGAAGATACAAATTTCTTTTGTGCTGGCT 1085
 QY 1223 TCAAGGAGAGTATGAGGAGCATCATGTTGTGCGAATCCAAGACAGAGCTCTTATAAATG 1282
 DB 1086 TAAAGAAACGATTTGAAATTCACCAAGCGGTGAATATTCATGACAAATGCTTTAGTTGCTG 1145
 QY 1283 CTGCTCAGCTGTCTGCTGTTACATAACTGTCGGCATTTACCGGATTAAGCAATTCGATT 1342
 DB 1146 CCGTACTTTATCCGATCGCTATATTAATGATCGTTTCTTACCAGATAAGAGCGATTGACT 1205
 QY 1343 TGGTTGATGAGGCTTTGTGCGAATGCTGAGAGTCAGCTTGATAGTCAACCTGAAGAGATTG 1402
 DB 1206 TAGTCGATGAGCAAGCGCAAGATTCGAGTGGAAATGAATTCGATGCCACAGAGCTCG 1265
 QY 1403 ATAACCTTTGAAGGAGGAGGATGAGCTGGAAATTTGAACCTTCACGCTTTGAAAGGAGA 1462
 DB 1266 ATCAAGTAAACGTCGGTTAATGCMACVAGAAATTTGAAGAGCGGCTTGAAAAAGAAAT 1325
 QY 1463 AGGATAAGCCAGCAAGCTCGACTTATAGAGTGGCGAAGAGCTTTGATGACCTGAGAG 1522
 DB 1326 CAGATGATGCAAGTAAAGAACCGGTTAGCCAACTTGCAAGAGAAATTTAGCAGATTTCGCTG 1395
 QY 1523 ACAAGCTTCAGCTCTCAGGATGAAATACAGAAAGGAGAAAGAGAGAAATTTGATGAGATTC 1582
 DB 1386 AAGAGCCAACTCAATGAAATGCAATGCGGAAACGGAAGAGAGAGTTAATCGGCTTT 1445
 QY 1583 GAAGCTTTAAGCAAGAAAGAGAGAGCTCATGTTTCTTTCGAGAGGCGAGAGCAAGAT 1642
 DB 1446 CCAATAAAGCGTGCAGAGATTGATAAAGCAAAACCAAGAAATTTAGAAGATGCTGAAATAAT 1505
 QY 1643 ATGACCTTGCAAGAGCTGCTGATCTTAGATATGCGCAATTTCAAGAGATGGAATCTCAA 1702
 DB 1506 ATGATTTAGAACGAGCTGCTGTTTGGGTCATGGAACAATTTCCGCAATTTAGACATGAAT 1565
 QY 1703 TTGCCCAACTTTGAAGGAACTTTCTTCTGAAGAGAAATGTGATGCTCACAGAAAACTGGGC 1762
 DB 1566 TGAAGATTTGAGAGAAAGAACGCTTAAGATTAAGTCAAAATGTTACAGAAATCGGTTA 1625
 QY 1763 CTGAACAC ---ATTGCTGAGGTTGTGAGCCCTTGGACAGGAAATTCAGTGTACGAGACTTG 1819
 DB 1626 CTGAACACGAAATTTGCGCAAGTGGTGGTCTTTAACTGGCAATTCCTGTGACAAAAATTAG 1685
 QY 1820 GCCAAATGAGAGGAGGTTGATTCGTTCTGCTGATGAGTTGATTCATAGCGGTTGTTGG 1879

Search completed: February 12, 2004, 17:47:36
 Elapsed time : 804 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:26:16 ; Search time 5756 Seconds
(without alignments)
13110.736 Million cell updates/sec

le: US-09-812-350-30

fect score: 3105

quence:

1 aagttatcaattttacaa.....tctttgggggtttttcata 3105

ring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

urched: 22781392 seqs, 12152238056 residues

al number of hits satisfying chosen parameters: 45562784

imum DB seq length: 0

imum DB seq length: 2000000000

-processing: Minimum Match 0%

Minimum Match 100%

Listing first 45 summaries

base :

EST:*

1: em_estba:*

2: em_esthm:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_man:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult NO.	Score	Query Match	Length	ID	Description
1	1268.6	40.9	3163	11	AY109444 Zea mays
2	605.4	19.5	633	9	AY826643
3	603.6	19.4	1015	28	AF106743
4	560	18.0	641	9	AU236881

ALIGNMENTS

RESULT 1
AY109444
LOCUS Zea mays CL1518_1 mRNA sequence.
DEFINITION Zea mays
ACCESSION AY109444
VERSION AY109444.1 GI:21213165
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3163)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S., Arthur,I.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3163)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES

source
1. 3163
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630374"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration with the overgo addressing of BACs in conjunction with the Maize Mapping Project"

626 a 846 c 942 g 490 t 259 others

GIN

Query Match 40.9%; Score 1268.6; DB 11; Length 3163;

Best Local Similarity 63.8%; Pred. No. 1.4e-279; Matches 1782; Conservative 0; Mismatches 992; Indels 18; Gaps 4;

141 GGATTACAAAGCTAATCGAAGATGAATCCAGAGAAATTCACACAGACAAACGAGAC 200
132 GGAATGACACAGAGACGAGCATGAATCCGGAACATTCACCCACAGACGACGAGGC 191
201 AATTGCTACAGCTCATGAGTAGCTGTGAATGACAGCATGCTCAATTCATCTTTTGA 260
192 GATCGTNNNNCCGACAGAGATTGCGTGGAGCGGCCACCGCGAGCTCACGCCGTGCA 251
261 TTTAGCTGGTGTGTGATCTCTGATCCACCGGTATATTCTCAAGCAATCTTAGTGC 320
252 CCGTCCCGAGTGTGGTGGCGACAGAGGGGGCATCTCGCGGAGGCGCATCNNNNNNN 311
321 CGGTGGCGAGAACGAGCTCAATCTGCT-----GAAAGAGTGATCAATCAAGCTTTGAA 374
312 NNN 371
375 GAAGTTCTTCAAAATCTCTCCACCTGATGATATTCACGCGAGTTCTAGCTTTATTA 434
372 GAAAGTGGCTCGAGTCCCGCGCCGCGACTCCGCTCCCGCGCTCCACGCGCTGATCAA 431
435 GGTCAATCGTGTGCTCAAGTGTCAAGTGTCAAGTGTCAATCTATTTGGCTGTGA 494
432 GGTCAATCGCGCGCGCGAGTCCGCGCAGAGAAACGCGGGGACTCGCACCTCGCGTGA 491
495 CCAGTTGATTATGGTCTTTTGAAGATTCTCAATCAGGGATTGTTGAACGAGTCGG 554
492 CCAGTGTGCTCGGCTGCTCGAGGACTCGAGATCTCCGACTGCTCAAGNNNNNN 551
555 TGATGAGCGGCGAGGGTAAAGTGTGAGTTGAGAGCTTCGTGGGAAGAGGAAGAA 614
552 NNN 611
615 AGTTGAGAGTGTCTCAGGCGACACAAATTTTCAAGCTTTAAAGACTTATGAGAGATT 674
612 NNNNNNNCCCGTNNNNNACCAATTTCCAGCGCTCAAGACATACGCGCGGACCT 671
675 GGTGAGCAAGCAGGAGAGCTGTGATCTGTGATTTGGTGTGATGAGGAGATTAGAAGT 734
672 CGTGAAGCAGGCGGAGAGCTGACCCCGTCAATCGCGCGCAGGAGAGATCCGCGCGT 731
735 CGTGAAGATTCTTCAGAGAGAACGAGAACAAATCTGTGCTTATTGAGAGCCAGGAGT 794
732 CGTGCATCTCTCGCGCGCAGCAACCAAGAACCAACCCGCTCTCATCGCGAGCCCGGCT 791
795 TGGTAAACAGCTGTGTTGAAGTTTACCAAAAGGATTGTGAAGAGAGATGTCGCCAA 854
792 TGGCAAGA CGGCGCTGTGAGAGGCGCTCCGCGAGCGCATCTGTTCCGCGGAGCTGCCCG 851
855 CAGTCTTACTGATGTGAGATTAAATTTCTGTTGACATGGGTGGTGTGTTGCTGTGTTAA 914

852 TAACCTCTCGAGCTCGCCTCATCGCCTCGACANNNNNNNNNNNNNNNNNNNNNNN 911
915 ATACCGAGAGAGTTTGAAGAAAGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGA 974
912 GTACCGCGCGAGTTTCGAGGAGCGGCTCAAGCCCGTGTCTCAAGGAGGTGGAAGAGCCGA 971
975 AGGCAAGTGAATCTCTTTATGATGAGATTCATTTGTTCTTGTGCTGGCAAACTGA 1034
972 GGGGAGGTCAATCTCTTCATCGCAGATACACTGCTCCNNNNNNNNNNNNNNNNNNNN 1031
1035 AGGGTCGATGATGAGCTTAATCTGTTCAAGCCCAATGTTAGCTAGAGGCGAGCTTCGATG 1094
1032 NNNNNCCATGAGCGCGGCAACTGTTCAAGCCAATGCTGGCGAGGACAGCTCAGGTG 1091
1095 CATTTGTGCTCAACGCTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCCCTTTGA 1154
1092 CATCGCGCCACCAACGCTGGAGGAGTACCGCAAGTACGTTGGAGAGGACGCGAGCTTCGA 1151
1155 GAGGAGTTCCAAACAAAGTCTATGTTGGGAGCCAAAGTGTGCTGACACCATTTAGTATCT 1214
1152 GCGGCGGTTCCAGCAGGTGTTGCTCGGAGCGGAGCTGCCCGACACCGTCAGCATTTCT 1211
1215 TAGAGACTCAAGGAGAGATATGAGGACATCATGTTGTGCGAAATCCAAAGACAGAGCTCT 1274
1212 GAGGGACTCAAGGAGAGATACGAGGGCACCATGGCTGAGGATCCAGGACCGGCGCT 1271
1275 TATAATGCTGCTCAGCTGTCTGCTGTTACATAACTGTTGCGGCAATTTACCGGATAAAGC 1334
1272 CGTGGTCCGCGGACAGCTCTCCGCGAGGTACATCATGGTGGTGGCGACCTGCCCTGACAAAGC 1331
1335 AATTGATTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGTAGTCAACCTGA 1394
1332 CATAGACTGTTGGAGAGGCTTGGCCCAATGTGAGGGTGCAGCTCGACAGCCAGCGGA 1391
1395 AGAGATTGATACTTTGAAAGAGAGATGAGCTGGAATTTGAACTTTCAGCCCTTGA 1454
1392 GAGATTGATACTTCGAGAGAGAGAAATCCAGCTTGAGGTTGAGCTCCACGCGCTGA 1451
1455 AAGGAGAGAGGATAAAGCCAGCAAGCTCGACTTATAGAGGTGCGGAAAGAGCTTGATGA 1514
1452 GAAGGAGAGAGGACAGGCCAGCAAGCCCGGCTGATTGAGGTGAGGAGAAATTTGACGA 1511
1515 CTTGAGAGCAAGCTTTGAGCTCTCAAGTGAATAACAGAAAGAGGAGAAAGAGAGAAATGA 1574
1512 TCTGAGGACAGCTCGACCCCTGACCATGAATANNNNNNNNNNNNNNNNNNNNNNNN 1571
1575 TGAGATTGAGAGCTTAAACAGAAAGAGAGAGCTCATGTTTCTTTTCAGAGGCGCAGA 1634
1572 TGAGATCAGGAAGCTGAAGCAGCGCGGAGGAGCTCCAGTTTCACTTCAGGAGGCGCGA 1631
1635 ACCAAGATATGACCTTTCAGAGAGCTGCTCATCTAAGATATGCGCAATTTCAAGAAAGTGA 1694
1632 GCCTCGATGAGCTGCTGCGCTGTGTCGCGACTCAAGTACGCGGCGCTCCAGGAATCGA 1691
1695 ATCTGCAATTTGCCCACTTTGAAGGAACTTTCTTGAAGAGAAATGTGATGCTCACAGAAA 1754
1692 CGCGCGATCTCCAAGCT---NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1748
1755 CGTTGGGCTGAAACATTTGCTGAGGTTGTGAGCCGTTGACAGGGATTCAGTGCAGGAG 1814
1749 CGTGGGCGCTGAAACAAATTTGAGAGGTGTTGAGCGCTTGAAGCGGTTATTCAGTGCACCG 1808
1815 ACTTGGCCAAATAGAGAGAGGTTGATTGTTCTTGTGATAGGTTGCATTAAGCGGCT 1874
1809 GCTTGGCCAGACGACAGAGAGGCTGTTGCGCTGCTGACAGGCTTCAACGAGGCT 1868
1875 TGTGGGACAGATCAAGCGGTAAATGACGTTTGTAGGCAATTTTAAGGTCAGGCGAG 1934
1869 GGTGCGCAGACAGAGGCTGTGAGCGCTGCGCAGAGCGGCTGCTGAGGTCAGGCGCGG 1928
1935 ACTTGTGAGGACAAACAGCCCACTGATCATTTTATTTCTTGGACCAACTGTTGTTGG 1994

LT 2
6643
'S
NITIO
SSION

327 CCTCAAGCAATCTCTAGTGTCCGGTGGCGAAGACGACGCTCAATCTGCTGTAAAGAGTGATC 386

361 AATCAAGCCTTGAAGAGCTTCTTCCATCTCTCTCCACCTGATGATATTCACGCGAGT 420

387 AATCAAGCCTTGAAGAGCTTCTTCCATCTCTCTCCACCTGATGATATTCACGCGAGT 446

421 TCTAGTCTTATTAAGGTCATTCGTCGTCAGCTCTCAGAAAGTCAGAGTGATCT 480
 447 TCTAGTCTTATTAAGGTCATTCGTCGTCAGCTCTCAGAAAGTCAGAGTGATCT 506
 481 CATTTGGCTGTTGACCAAGTGAATATGGGCTCTCTTGAAGATCTCAAAATCAGGATTTG 540
 507 CATTTGGCTGTTGACCAAGTGAATATGGGCTCTCTTGAAGATCTCAAAATCAGGATTTG 566
 541 TTGAACCAAGTCGCTAGCGACGCGGAGGTAAGTCTCAGGTTGAGAAAGTCTCGTGGG 600
 567 TTGAACCAAGTCGCTAGCGACGCGGAGGTAAGTCTCAGGTTGAGAAAGTCTCGTGGG 626
 601 AAAGAAG 607
 627 AAAGAAG 633

JLT 3
 US 06743/c 1015 bp DNA linear GSS 06-NOV-2000
 AFI06743 Arabidopsis thaliana Col-0 Arabidopsis thaliana genomic
 clone 060-2E1.Km, genomic survey sequence.
 AFI06743
 AFI06743.1 GI:4588414

GSS
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

Koncz, C., Szabados, L., Grunberg, B. and Schaefer, S.
 Gene identification with sequenced T-DNA tags in Arabidopsis
 Unpublished

Contact: Koncz C
 Abteilung Genetische Grundlagen der Pflanzenzuechtung
 Max-Planck Institut fuer Zuechtungsforchung
 Carl von Linné weg 10, Cologne, D-50829, Germany
 Email: koncz@mpiz-koeln.mpg.de
 Line 060-2 was obtained by transformation with the T-DNA of
 pPCV6NFHYg Agrobacterium binary vector; the right border junction
 of T-DNA insertion 060-2E1 was isolated by PCR amplification using
 the BH2 (5'- CGTATGTTTATCGCACTTGCATCGG-3') and Km2 (5'-
 CAGTCATAGCGAATAGCTCTCCACC-3') primers and sequenced with the BH2
 and Km2 primers as described by Mathur et al. Plant J. (1998) 13,
 707-716; the PCR-amplified DNA fragment of 1.05 kb extends from an
 EcoRI site to the right border junction of pPCV6NFHYg T-DNA tag,
 sequences of the right T-DNA border are excluded from the
 submission

Class: transposon-tagged.

TURES

source

1. .1015
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Col-0"
 /db_xref="taxon:3702"
 /clone="060-2E1.Km"
 /clone_lib="Arabidopsis thaliana Col-0"
 302 a 225 c 198 g 288 t 2 others

E COUNT
 GIN

very Match 19.4%; Score 603.6; DB 28; Length 1015;
 est Local Similarity 86.9%; Pred.No. 2.6e-127;
 atches 730; Conservative 0; Mismatches 6; Indels 104; Gaps 2;

1 AAAGTTATCAATTTTACCAATTAACCGCTATAATCTGTTGATTTCTGCAAAAGAGAA 60
 886 AAAGTTATCAATTTTACCAATTAACCGCTATAATCTGTTGATTTCTGCAAAAGAGAA 827
 61 GACTTTTACCAGAGAGGTCCTCTGGCTCATTTGAAGAAACTCAAGCAACAACCCAG 120

Db 826 GACTTTTACCAGAGAGGTCCTCTGGCTCATTTGAAGAAACTCAAGCAACAACCCAG 767
 QY 121 TTCTCATATATCGTTTTAA----- 139
 Db 766 TTCTCATATATCGTTTTAAAGGTAATAATGATCGCAATCTTTGTTCTCATTGTTGTTT 707
 QY 140 ----- 139
 Db 706 TGTGTTTGTGATTAGGTTTACAAAGATACATGATTAGTTTTTTTTTTTTTTTTTTT 647
 QY 140 --GGATTACAAAAGCT--AATCGAAGATGAATCCAGAGAAATTCACACACAAGCAAAACG 196
 Db 646 CGGGGTTTACAAAAGCTAAATCGAAGNTGAATCCGNGAAATTCACACACAAGCAAAACG 587
 QY 197 AGACAATTTGCTACAGCTCATGAGCTAGCTGTGATGAGGACATGCTCAATTCACCTCCT 256
 Db 586 AGACAATTTGTAACAGCTCATGAGCTAGCTGTGATGAGGACATGCTCAATTCACCTCCT 527
 QY 257 TGCATTTAGCTGCTGCTTTGATCTCTGATCCACCGGTATATTTCCCTCAAGCAATCTCTA 316
 Db 526 TGCATTTAGCTGCTGCTTTGATCTCTGATCCACCGGTATATTTCCCTCAAGCAATCTCTA 467
 QY 317 GTGCCGTTGGCGAGAACCGAGCTCAATCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGA 376
 Db 466 GTGCCGTTGGCGAGAACCGAGCTCAATCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGA 407
 QY 377 AGCTTCCCTTCAATCTCTCCACCTGATGATATTTCCAGCGAGTTCTAGTCTTATTAAGG 436
 Db 406 AGCTTCCCTTCAATCTCTCCACCTGATGATATTTCCAGCGAGTTCTAGTCTTATTAAGG 347
 QY 437 TCATTTGCTGCTGCTCAAGCTGCTCAGAAAGTCACGAGTGATATACTCATTTGGCTGTTGACC 496
 Db 346 TCATTTGCTGCTGCTCAAGCTGCTCAGAAAGTCACGAGTGATATACTCATTTGGCTGTTGACC 287
 QY 497 AGTTGATTATGCTGCTCTTGAAGATTTCTCAAAATCAGGAGTTTGTGACCAAGTCGGTG 556
 Db 286 AGTTGATTATGCTGCTCTTGAAGATTTCTCAAAATCAGGAGTTTGTGACCAAGTCGGTG 227
 QY 557 TAGCGACGCGGAGGTTGAAGTCTGAGTTGAGAGCTTCGTGGGAAAGAGGAGGAAG 616
 Db 226 TAGCGACGCGGAGGTTGAAGTCTGAGTTGAGAGCTTCGTGGGAAAGAGGAGGAAG 167
 QY 617 TTGAGAGTGCTTTCAAGGTCACCAAAATTTTCAAGCTTTTAAAGACTTATGGAAGAGATTGG 676
 Db 166 TTGAGAGTGCTTTCAAGGTCACCAAAATTTTCAAGCTTTTAAAGACTTATGGAAGAGATTGG 107
 QY 677 TTGAGACAGCAGGAGGAGCTTGATCTCTGATGTTGCTGCTGATGAGGAGATTAGAGAGTCG 736
 Db 106 TTGAGCAACGACGAGGAGCTTGATCTCTGATGTTGCTGCTGATGAGGAGATTAGAGAGTCG 47

RESULT 4

AU236881

LOCUS

DEFINITION

RNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 641)

Seiki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,

Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J.,

Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu

, M., Hayashizaki, Y. and Shinozaki, K.

Large scale analysis of Arabidopsis full-length cDNA

Unpublished

CONTACT: Motoaki Seki

Plant Functional Genomics Research Group

COMMENT

RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: maeiki@rc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda phage vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified phage vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

URES Location/Qualifiers
 source
 1. .641
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAF15-30-A07"
 /tissue_type="mixture of silique and flower"
 /lab_host="DH10B"
 /clone_lib="RAF15"
 /note="Site 1: BamHI; Site 2: SalI"
 { COUNT 190 a 146 c 128 g 174 t 3 others
 IN

ery Match 18.0%; Score 560; DB 9; Length 641;
 est Local Similarity 99.3%; Pred. No. 2.2e-117;
 tches 571; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

1 AAAGTTATCAATTTTAAACATACCGCTATATCTGTTGATCTCTGCAAAAGAGAA 60
|||||
61 AAAGTTATCAATTTTAAACATACCGCTATATCTGTTGATCTCTGCAAAAGAGAA 120
|||||
61 GACTTTTACCAGAGAAAGTCTCTGGCTCATTTGAAGAACTCAACGAAACAAACCCAG 120
|||||
121 GACTTTTACCAGAGAAAGTCTCTGGCTCATTTGAAGAACTCAACGAAACAAACCCAG 180
|||||
121 TTCTATATATGTTTTAAGGATTAACAAAGCTAATCGAAGATGATCCAGAAATTC 180
|||||
181 TTCTATATATGTTTTAAGGATTAACAAAGCTAATCGAAGATGATCCAGAAATTC 240
|||||
181 ACACACAGACAAACGAGCAATTCGTCAGCTCATGAGCTAGCTGATGATCGAGCAT 240
|||||
241 ACACACAGACAAACGAGCAATTCGTCAGCTCATGAGCTAGCTGATGATCGAGCAT 300
|||||
241 GCTCAATTCACCTTTGATTTAGCTGATTTGATCTGATCCCAACCGGTATATT 300
|||||
301 GCTCAATTCACCTTTGATTTAGCTGATTTGATCTGATCCCAACCGGTATATT 360
|||||
301 CCTCAAGCAATCTAGTCCGGTGGCGGAGACGAGCTCAATCTGCTGAAGAGTGATC 360
|||||
361 CCTCAAGCAATCTAGTCCGGTGGCGGAGACGAGCTCAATCTGCTGAAGAGTGATC 420
|||||
361 AATCAAGCTTTGAAGAGCTTCTTCAACATCTCTCCACCTGATGATATCCAGCGAT 420
|||||
421 AATCAAGCTTTGAAGAGCTTCTTCAACATCTCTCCACCTGATGATATCCAGCGAT 480
|||||
421 TCTAGTCTTTAAGGTCATTCGTCGCTCAAGCTGCTCAGAGTCAAGGTCGATCT 480
|||||
481 TCTAGTCTTTAAGGTCATTCGTCGCTCAAGCTGCTCAGAGTCAAGGTCGATCT 540
|||||
481 CATTCGCTGTTGACCAAGTGTATTAT-GGGCTCTTTGAAGATTCCTCAATCAGGATTT 539
|||||
541 CATTCGCTGTTGACCAAGTGTATTAT-GGGCTCTTTGAAGATTCCTCAATCAGGATTT 600
|||||
540 GTTGAACGAGTCCGTTGATCGGACGCGGAGGTAA 574
|||||
601 GTTGAACGAGTCCGTTGATCGGACGCGGAGGTAA 635

```

CB894504 846 bp mRNA linear EST 24-APR-2003

DEFINITION

EST647296 HOGA Medicago truncatula cDNA clone HOGA-31M8, mRNA

sequence.

ACCESSION CB894504

VERSION CB894504.1 GI:30101673

KEYWORDS EST.

SOURCE Medicago truncatula (barrel medic)

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroside I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

1 (bases 1 to 846)

Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,

Utterback, T., Cho, J., and Fraser, C.M.

ESTs from roots of Medicago truncatula treated with

oligogalacturonides of DP 6-20

Unpublished

Contact: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMD76TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtc gat CC).

Location/Qualifiers

1. 846

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="HOGA-31M8"

/tissue_type="3 day old seedling roots"

/dev_stage="24 hours after treatment in the dark at 26 C

with 0.5 mg/ml oligogalacturonides (DP 6-20) in the

presence of 100 ug/ml Gentamicin

/lab_host="XLOLR"

/clone_lib="HOGA"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Unizap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

helper phage and propagated in SOLR cells."

BASE COUNT 270 a 139 c 243 g 194 t

ORIGIN

Query Match 17.5%; Score 543.4; DB 14; Length 846;

Best Local Similarity 78.7%; Pred. No. 1.6e-113;

Matches 665; Conservative 0; Mismatches 171; Indels 9; Gaps 1;

Qy 1327 GATTAACCAATTTGTTTGGTTGATGAGCTTGCCGATGTGAGCTCCAGCTTGATGT 1386

Db 1 GACAAGGCCATTTGATTTAGTTGATGAGGCGCATGTGTAATGTGAGAGTTCACTTGATGT 60

Qy 1387 CAACCTGAAGATTGATTAACCTTGAAGGAGGAGGATGCGCTGGAAATTTGAATTCAC 1446

Db 61 CAACCTGAGGAATTTGCAATCTTTGAAGGAGGAGGATGCGANTTAGAGTTGAATTCGAT 120

Qy 1447 GCCTTGGAAAGGGAGGAGGATTAAGCCAGCAGCAAGCTTCGACTTATAGAGGTGCGGAAGAG 1506

Db 121 GCTCTAGAGAAAGAGGAGGAGCAAGCTAGCAAGCCGCTCTCTGTAGATGTACCGAGAGAA 180

Qy 1507 CTTGATGACCTTGAGAGCAAGCTTCAGCCTCTCAGATGAAATACAGAGGAGAGAGAG 1566

Db 181 CTTGATGACTTGAGAGCAAGCTTCACCTCTCGAAGTGAAGTATAGTAAAGAGAGAGAG 240

Qy 1567 AGAATTCATGAGATTTCGAAGGCTTAAACAGAAAGAGAGAGAGCTCATGTTTCTTTTCGAG 1626

Db 241 AGGATTGATGAGATTTCGAAGGCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300

1627 GAGCGAGAACGAGATATGACTTTCAGAGCTGCTGATCTTAAGATATGGCGCAATTCAA 1686
 301 GAGCGAGAGAGCGGTATGATCTAGCAGAGCTGCGACCTGCGATATGGTGCATTTGAA 360
 1687 GAAGTGGAACTGCAATTTGCCCAACTTTGAAGAACTTCTT-----TCTGAAGAGAAAT 1737
 361 GAGTGGAACTGCAATTTAAAAATCTTGAAGTAGCACTGATGGGAACACGGATGAGAAC 420
 1738 GTGATGCTCAGAGAAAGCTTGGCGCTGMAACATCTGCTGAGGTTGAGCCGTTGACA 1797
 421 TTAATGTTGACGGAACAGATTGGACCGGACCAATAGCCGAGGTGTTAGCCGATGACT 480
 1798 GGGATTCAGTGCAGCAGACTTGGCCAAATGAGAAGAGAGGTTGATGTTCTTCTGCTGAT 1857
 481 GGTATACAGTACTAGCTCGGCCAAATGAGAAAGCAAGTTGGTGGACTTGGTGAC 540
 1858 AGGTTGCATAGAGCGGTTGGGACAGAACTCAAGCGGTAAATGACGTTCTTGAGGCAATT 1917
 541 AGACTGCACACTAGAGTGGTGGGACAGACCAAGCGGTTAATGCTGTTGCTGAGGCTGTA 600
 1918 CTAAGTCAAGGCGAGGACTTGTAGGGCAACACAGCCAACTGGATCATTTCTTATTCCTT 1977
 601 TTGAGATCAAGAGCTGTTTGAAGAGACCCCAACCACTGGTTCTTCTTATTCCTT 660
 1978 GGACCAACTGTTGTCGAACTGAGCTGCCAAGGCTTCTTCTGAGCAGCTGTTTGTAT 2037
 661 GGTCCGACTGTTGTTGTAAGTCTGAGCTTGCAGAGGCTTCTTGCAGAGGCTATTTGAT 720
 2038 GATGAACCTCTTAGTTCGATGATGATGTCGGAATATATGGAACAACTCTGCTCT 2097
 721 GATGAACCTCTTAGTTCGATGATGATGTCGGAATATATGGAACAACTCTGCTCT 780
 2098 CGCTCATTTGGGCAACCACTGAGGCTATGCTTCCACGAGGAAGTGCACAACTAATCTGAG 2157
 781 AGATGATGTTGTCACCCAGGCTATGTTGGACATGAGGAAGGGGTCATTTAACGAA 840
 2158 GCTGT 2162
 841 GCTGT 845

ULT 6
 50711
 US
 INITIATION
 BOGAQ17TF BOGA Brassica oleracea genomic clone BOGAQ17, genomic
 survey sequence.
 BH250711
 BH250711.1 GI:17074863
 GSS.
 Brassica oleracea
 ORGANISM
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 818)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1. 818
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"

TURES
 source

/db_xref="taxon:3712"
 /clone="BOGAQ17"
 /clone_lib="BOGA"
 /note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHS01 using BstXI linkers"
 BASE COUNT 222 a 158 c 217 g 221 t
 ORIGIN
 Query Match 17.5%; Score 542; DB 28; Length 818;
 Best Local Similarity 85.7%; Pred. No. 3.2e-113;
 Matches 615; Conservative 0; Mismatches 100; Indels 3; Gaps 1;
 1439 AACTTCACGCCCTTGGAAAGGAGAGATTAAGCAGCAAGAGCTCGACTTATAGAGGTGC 1498
 77 AACTTCTCTTTTGGGCAAGTCTCTGATATTGTTATCTTGAATTAATGATCTGTAGGTGC 136
 1499 GGAAGAGCTTATGACCTGAGAGACAAGCTTCAAGCTTCAAGATGAATACAGAAAGG 1558
 137 GGAAGAGCTTATGACCTGAGGGAACAAGCTTCAAGCTTCAAGATGAATACAGAAAGG 196
 1559 AGAAGAGAGAAATTCATGAGATTTCGAAGGCTTTAAACAGAAAGAGAGCTCATGTTTT 1618
 197 AGAAGAGAGAAATTCATGAGATTTCGAAGGCTTTAAACAGAAAGAGAGCTCATGTTTT 256
 1619 CTTTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 1678
 257 CTTTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 316
 1679 CAATTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 1738
 317 CAATTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 373
 1739 TGATGCTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 1798
 374 TGATGCTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 433
 1799 GATTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 1858
 434 GATTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 493
 1859 GATTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 1918
 494 GATTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 553
 1919 TAAGTTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 1978
 554 TAAGTTCAGAGGAGCAGACGAGATATGACCTTTCAGAGCTGCTGATCTAAGATATGCG 613
 1979 GACCAACTGCTGTTGGCAAACTGAGCTGCGCAAGGCTTTCAGAGCTGCTGATCTAAG 2038
 614 GACCAACTGCTGTTGGCAAACTGAGCTGCGCAAGGCTTTCAGAGCTGCTGATCTAAG 673
 2039 ATGAAACTCTTTAGTTCGATGATGATGTCGGAATATATGGAACCAACTCTGCTCTC 2098
 674 ATGAAACTCTTTAGTTCGATGATGATGTCGGAATATATGGAACCAACTCTGCTCTC 733
 2099 GCTCATTTGGGCAACCACTGAGGCTGTTGGTTCAGAGGAGGTTGCAACTTAAGTGA 2156
 734 GCTCATTTGGGCAACCACTGAGGCTGTTGGTTCAGAGGAGGTTGCAACTTAAGTGA 791

RESULT 7
 BH465756/c
 LOCUS
 DEFINITION BOGHI04TF BOGE Brassica oleracea genomic clone BOGEI04, genomic
 survey sequence.
 ACCESSION BH465756
 VERSION BH465756.1 GI:17664770
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eucosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 863)
 Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of *Brassica oleracea*
 Unpublished
 Other_GSSs: BOGEI04TR
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

URES
 source

1..863
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGEI04"
 /clone_lib="BOGE"

/note="Vector: pBstXI; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pBstXI using BstXI linkers"

COUNT 213 a 234 c 165 g 251 t

Query Match 17.4%; Score 541.4; DB 28; Length 863;
 Best Local Similarity 87.2%; Pred. No. 4.5e-113;
 Matches 606; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

1439 AACTTCACGCCCTTGAAAGGGAGAGGATAAGCCAGCAAGCTCGACTTATAGAGTGC 1498
 697 AACTTCCTCTTTGGGCAAGTCTCTGATTTGTTATCTGATTTATGATCTGATGTC 638

1499 GGAAGAGCTTGATGACCTGAGAGCAAGCTTCAGCCTCTCACGATGAATACAGAAAGG 1558
 637 GGAAGAGCTTGATGACCTGAGGAGCAAGCTACAGCCTCTCAGATGAAGTACAAAAGG 578

1559 AGAAGAGAGATTGATGATTCGAAGCTTAACAGAGCAAGAGAGAGCTCATGTTT 1618
 577 AGAAGAGAGATTGATGATTCGACGCTTAAACAGAAAGAGAGAGAGCTCATGTTT 518

1619 CTTTCAGAGAGCAAGCAAGATATGATCTGCAAGAGCTGCTGATCTAAGATATGCG 1678
 517 CTTTCAGAGAGCAAGCAAGATATGATCTGCAAGAGCTGCTGATCTAAGATACGAG 458

1679 CAATTCAAGAGTGAATCTGCAATTGCCCACTTGAAGGAACTTCTTCTGAAGAGATG 1738
 457 CAATCCAAGAGTGAATCTGCAATAGCCCACTCGAAGGAA---CTTCTGACGAGAACA 401

1739 TGATCTCACAGAAACCTTGGCTGAGCAATGCTGAGGTTGTGAGCCGTTGGACAG 1798
 400 TGATCTCACAGAAACCTTGGCTGAGCAATGCTGAGGTTGTGAGCCGTTGGACAG 341

1799 GGATTCAGAGTGAATCTGCAATTGCCCACTTGAAGGAACTTCTTCTGAGAGATG 1858
 340 GGATTCAGAGTGAATCTGCAATTGCCCACTTGAAGGAACTTCTTCTGAGAGATG 281

1859 GGTTCATTAAGCGGTTTGGGACAGATCAAGCGGTAATGCGAGTTTCTGAGGCAATTC 1918
 280 GGTTCATTAAGCGGTTTGGGACAGATCAAGCGGTAATGCGAGTTTCTGAGGCAATTC 221

1919 TAAGTTCAGGCGAGGCTTGTAGGCGACACACCACTGAGATCATCTTATCTCTTG 1978
 220 TAAGTTCAGGCGAGGCTTGTAGGCGACACACCACTGAGATCATCTTATCTCTTG 161

1979 GACCAACTGGTTCGCAAAACTGAGCTCGCAAGGCTTCTGCTGAGAGCTGTTTGTATG 2038
 160 GCCCAACTGGTTCGCAAAACTGAGCTCGCAAGGCTTCTGCTGAGAGCTGTTTGTATG 101

2039 ATGAATAAACCCTTAGTTCGGATTGATGTCGGAAATATATGGAACAACACTCTCTCTC 2098

Db 100 ATGAAACCTCTTAGTTGCGATTGATATGCGAGTATATGGACAGCACTCTGTCTCTC 41
 QY 2099 GCCTCATTTGGGCGACACCAAGGGTATGTTGGTCAAC 2133
 Db 40 GCCTCATTTGGGCGACACCAAGGGTATGTTGGTCAAC 6

RESULT 8
 LOCUS CB894478
 DEFINITION EST647270 HOGA Medicago truncatula cDNA clone HOGA-31G20, mRNA
 ACCESSION CB894478
 VERSION CB894478.1 GI:30101647
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)

ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eucosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 837)
 AUTHORS Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
 Uterback, T., Cho, J. and Fraser, C.M.
 TITLE ESTs from roots of *Medicago truncatula* treated with
 oligogalacturonides of DP 6-20
 JOURNAL Unpublished
 COMMENT Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMD746TK
 More information is available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gtc gat CC).
 FEATURES
 Location/Qualifiers

1..837
 source
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="HOGA-31G20"
 /tissue_type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C
 presence of 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_host="XLOLR"
 /clone_lib="HOGA"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in SOLR cells."

BASE COUNT 258 a 147 c 236 g 196 t
 ORIGIN
 Query Match 17.0%; Score 529; DB 14; Length 837;
 Best Local Similarity 78.0%; Pred. No. 3.1e-110;
 Matches 653; Conservative 0; Mismatches 175; Indels 9; Gaps 1;

1496 TCGGAAAGAGCTTGATGATGATTCGAGGCTTAACAGAAAGAGAGAGCTCATGT 1555
 Db 1 TACGGAGAGAACTTGATGATGATTCGAGGCTTAACAGAAAGAGAGAGCTCATGT 60

1556 AGGAGAAAGAGAGAACTTGATGATGATTCGAGGCTTAACAGAAAGAGAGAGCTCATGT 1615
 Db 61 AAGAGAAAGAGAGAACTTGATGATGATTCGAGGCTTAACAGAAAGAGAGAGCTCATGT 120

1616 TTTCTTTGAGGAGGAGAGATATGACCTTTCAGAGAGCTGCTGATCTAAGATATG 1675
 121 TTGCACTACAGGAGGAGAGGCGGTATGATCTAGCAAGAGCTGCGACCTCGATATG 180
 1676 GCGCAATTCAGAGAGTGGAAATCTGCAATCCCAATCTGGAAGCACTTCT-----T 1726
 181 GTGCAATTCAGAGAGTGGAAATCTGCAATTCCTGGAAGCACTTCTGGAAGCACTTCT 240
 1727 CTGAGAGAGATGATCTCACAGAGAGAGTGGCTGCGCTGAGCAATCTGCTGAGGTGTA 1786
 241 CGGATGAGAGCTTAAATGTTGACGGAAGAGTGGCCGAGCAAAATAGCCGAGGTGTTA 300
 1787 GCCCTTGGACAGGAGATCCAGTACAGAGAGTGGCCCAAAATAGAGAGAGGTGATTTG 1846
 301 GCCGATGAGCTGTTATACCACTGAGTGGCTCGGCCCAAAATAGAGAGAGCAAGTGGTTG 360
 1847 GTCTTCTGATAGTTCATAGCGGGTGTGGGACAGATCAAGCGGTAAATGCGAGTTT 1906
 361 GACTTGTGTGACAGACTGCACACTAGAGTGTGGGACAGAGCAAGCGGTAAATGCTGTTG 420
 1907 CTGAGGCAATTCCTAAGGTCAAGGCGAGAGCTTGGTAGGCGACCAACAGCAACTGGATCAT 1966
 421 CTGAGGCTGTTATGAGATCAAGAGCTGTTTAGAGAGAGCCCAACCAACTGGTTCT 480
 1967 TCTTATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTGCGCAAGGCTCTTCTGAGC 2026
 481 TCTTATTCCTTGGTCCGAGCTGTTGTTGTAAGAGCTGAGCTTGCAGAGGCTCTTGCAGAGC 540
 2027 AGCTGTTGATGATGAAGAACTCTTAGTTCGGATGATGATGTCGGAATATATGGAACAC 2086
 541 AGCTATTTGATGATGAAGAACTCAGTTGGTGAAGATTTGATGTCGGAATATGGAACAC 600
 2087 ACTCTGCTCTCGCTCATTTGGGCGACACAGAGGTATGTTGGTCAAGAGAGGTGGAC 2146
 601 ACTCTGATCAAGATGATGTTGGTCAAGAGGTATGTTGGTCAAGAGAGGTGGAC 2206
 2147 AACTAATGAGGCTGTGAGAGGCGAATTTGTTGATCTCTTTGATGAAGTGGAGA 2266
 661 AATTAACAGAGCTGTGAGAGAGGCGATACAGTGTGTTGTTGATGAAGTGGAGA 720
 2207 AGGCTCATGTTGCTCTTCAACACTGCTGCAAGTTCGAGTTCGATGATGCTGATGACAG 2266
 721 AAGCACACATCTGTTATTCACACTCTTCTCCAAAGTCTTGGATGATGAGGAGTGTGACTG 780
 2267 ACGGCGAGGCGAGCAGCTGATTTCAAGAGCTCGGTGATATCATGACATCAAAACC 2323
 781 ACGGACAGGCGAGCAGCTGATTTTAGAAGACTGATCATCATGACCTCTTAACC 837

ULT 9
 49719
 US
 INITIATION
 CB349719 843 bp mRNA linear EST 10-APR-2003
 vinifera cDNA clone CAB28G0007_IIIaF_All 5', mRNA sequence.
 SESSION
 CB349719.1 GI:28970686
 EST.
 WORDS
 Vitis vinifera
 Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; Vitaceae; Vitis.
 1 (bases 1 to 843)
 Goes da Silva, F., Iandolo, A., Lim, H., Baek, J., Jones, K. and Cook
 D.
 Expressed sequence tags from cabernet sauvignon berries at various
 developmental stages
 Unpublished
 Contact: Douglas Cook, PhD
 CAES Genome Facility
 UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561

Fax: 530 754 6617
 Email: drcook@ucdavis.edu
 Seq primer: ACGGTACGGACATATGCC.
 Location/Qualifiers
 1..843
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultivar="Cabernet Sauvignon"
 /db_xref="taxon:29760"
 /clone="CAB28G0007_IIIaF_All"
 /sex="Hermaphrodite"
 /dev_stage="Veraison"
 /lab_host="DH5alpha"
 /clone_lib="Cabernet Sauvignon Berry - CAB28G"
 /note="Organ: Berry; Vector: pDNR; Site 1: SfiI; Site 2:
 SfiI; CAB28G is a cDNA library of Cabernet Sauvignon Clone
 8 berries. Samples were collected at veraison (transition
 between stage II and stage III of berry growth) from
 field-grown vines 60 days after full bloom. Berries were
 still green but soft. Sampled vines were located at the
 University of California, Davis, experimental vineyard.
 cDNAs were made by oligo-dT priming and directionally
 cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGTATCAACGAGAGTGGCATTACGCCGG-3' and
 5'-ATTCTAGAGCGGAGCGCGGACATG-dt(30)NN-3'. Library was
 constructed using the Clontech Creator SMART kit and
 size-selected to contain the 0.5-3 kb size fraction."
 BASE COUNT 218 a 166 c 251 g 207 t 1 others
 ORIGIN
 Query Match 16.3%; Score 506.4; DB 14; Length 843;
 Best Local Similarity 75.7%; Pred. No. 4.7e-105;
 Matches 640; Conservative 0; Mismatches 202; Indels 3; Gaps 1;
 QY 1654 AGAGCTGCTGATCTAAGATATGCGCAATTCAGAGAGTGGATCTGCAATTTGCCAATCTT 1713
 Db 2 AGAGCTGCTGATTTGCGATATGCGCAATTCAGAGAGTGGAGGCTGCTATAGCAAACTT 61
 QY 1714 GAAGGAATCTTCTTCAAGAGAGATGTGATGCTCAAGAGAAAGCTTTGGGCTTGAACATTT 1773
 Db 62 GAAGGA---CCACTGATGAGAACATGATGTTAAGGAGACTGTTGGGCTTGAACATTT 118
 QY 1774 GCTGAGTGTGTGAGCGGTGGACAGGAGATTCAGTGCAGAGACTTTGGCCAAATGAGAG 1833
 Db 119 GCTGAGTGTGTGAGCGGTGGACAGGAGATTCAGTGCAGAGACTTTGGCCAAATGAGAG 178
 QY 1834 GAGAGTGTGATGTTGCTGCTGATAGTGTGATAGCGGTTGTCGAGAGAGATCAAGAG 1893
 Db 179 GAAGGTTAATTTGGGCTTGGCGAGAGGCTGCAATCAGAGAGTGTGGGACAGATCAGGCA 238
 QY 1894 GTAAATGCAAGTTTCTGAGGCAATTCCTAAGGTCAGAGGAGGAGCTTTGGTAGGCAACAG 1953
 Db 239 GTCAGTGTGTGGCAGAGGCTGCTTTGAGTTCAGAGGCTGGTGTAGGAAGGCAACAA 298
 QY 1954 CCAGCTGATCATCTTATCTTGGACCAACTGCTGTTGGCAAACTGAGCTGCCAAG 2013
 Db 299 CCAGCTGTTCTGTTCTTCTTCTTGGGTCCTGCTGTTGGTGAAGACTGAGCTGCCAAG 358
 QY 2014 GCTTCTTGTGAGCAGCTGTTTGTGATGATGAAGAACTCTTGTGTTGGATGATATGTCGAA 2073
 Db 359 GCGCTGGCTGAACAGCTCTTGTGATGATGAAGAACTCTTGTGTTGGATGATGATGTCGAG 418
 QY 2074 TATATGGAACAACTCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2133
 Db 419 TATATGAGCAGCAATTCAGTTTCAAGATTTGATTGGGCTCTCTCTGCTCTGCTCTGCTCTGCT 478
 QY 2134 GAGGAGGTGGACAACTTAATCTGAGGCTGTGAGGAGGAGGAGCTTATTTGTCTATCTCTTT 2193
 Db 479 GATGAGGTGGGCAACTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 538
 QY 2194 GATGAGGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2253
 Db 539 GATGAGGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 598

FEATURES

source

2254 GGTGATTGACAGAGCGGACAGGACAGTCTGATTCAGGACTCGGTGATATCATG 2313
 599 GGAAGGTTAACTGATGTCACAGGCGGACACTGTAGACTTCACAAACACAGTGATCATATG 658
 2314 ACATCAAACTTGGTGTCTGACACACTCTCTTCGAGGCTAACTGGGAAAGTAAACAATGGAA 2373
 659 ACTTCAAACTTGGAGCTGAGCATCTCTCTCGGGTTTAGTAGGCAAGTGTACAAATGCAA 718
 2374 GTGCCCGGAGCTGTGTGATCGGGAGGTGAGGAAACACTTCAGACAGAGCTCTTGAAC 2433
 719 GATGCTCGTGACCGGGTGTATGACAGAGGTGAGNAGGCACTTCAGGCCAGAGTTGCTTAAAC 778
 2434 AGGCTTCAACAGAGATTGTGTGCTTGCACCCCTTTTACATGACCAAGTTGAGGAAAGTAGCT 2493
 779 CGGTGATGATGAATGTGTGATTTGATTTGATCCCTCTCATATCATGATCACTGAGAAAGTTGCC 838
 2494 CGGCT 2498
 839 AGGCT 843

LOC 10
 84385

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

ORIGIN

Query Match 16.3%; Score 505.8; DB 10; Length 736;
 Best Local Similarity 81.3%; Pred. No. 6.2e-105;
 Matches 599; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 451 CAAGCTGCTCAGAGTTCAGAGGTGATCTACTATTTGGCTGTGTGACCACTGATTTATGGGT 510
 Db 1 CAAGCTGCTCAGAA-TTACGTGTGATCTACTATTTAGCTGTGTGATCAATCTTGGGA 59
 QY 511 CTCTTTGAAGATTCTCAATCAGGGATTGTTGAAACGAGTCCGTGTAGCCGACGCGGAGG 570
 Db 60 ATTCTTGAAGATTCTCAGATTGCTGATTGTTTAAAGAGCTGTGTGCTTTCTAGG 119
 QY 571 GTAAAGTCTGAGGTGTGAGAGCTTCGTGGGAAAGGAAAGGAAAGGTTGAGAGTCTTCA 630
 Db 120 GTTAAAGTCTGAGGTGTGAAAGCTGAGGGGTAAAGATGGGAAAGGTTGAAAGTCTTCC 179
 QY 631 GGGGACACAAATTTTCAAGCTTTAAAGACTTATGCAAGAGATTGTTGTTGACGACGAGG 690
 Db 180 GGTGATACGAATTTTCAAGCATTTGAAGACTTATGTAGAGATTGTTGTTGAAACAGAGG 239
 QY 691 AAGCTTGATCTGTGATTTGGTGTGATGAGGAGATTAGAAGAGTCTGAGGATCTTTTCG 750
 Db 240 AAGCTTGATCTGTGATTTGGGACCGGACGAGAGATTTCGAGAGTTGTCAGGATTTTCA 299
 QY 751 AGGAGACGAGAAACATCTCTGTGCTTATTGAGAGCCAGAGTTGGTAAACAGCTGTG 810
 Db 300 AGGAGGACTTAAGATTAATCCAGTTCTTATTGTGTGAGCTGTGTGGGAAAGTCTGCT 359
 QY 811 GTTGAAGGTTTGAACAAAGGATTGTTGAAGGAGATGTCGCCAAGTCTTTACTGATGTG 870
 Db 360 GTTGAAGGTTTGGCTCAGAGATTGTTAGAGGTGATGTTCTGCAATCTTGTGATGTT 419
 QY 871 AGATTAAATTTGCTGTGACATGGGTGCTGTAGTGTGCTGTGCTGTGCTGTGCTGTGCT 930
 Db 420 AGTTAAATTTGCTGTGATGAGGAGTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 479
 QY 931 GAAGAAAGGTTGAATCTGTTTGAAGAGTGTGAGGAGCTGTGAGGAAAGTGTGATCTC 990
 Db 480 GAAGAGAGGTTGAAGGCTGTGTTTGAAGAGTGTGAGGAGCTGTGAGGAAAGTGTGATCT 539
 QY 991 TTATTGATGAGATTCAATTTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1050
 Db 540 TTCATGATGAGATTCACTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 599
 QY 1051 GCTAATCTTTTAAAGCCCAATGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1110
 Db 600 GCTAATCTTTTAAAGCCCAATGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 659
 QY 1111 CTGGAAGATACAGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTCCACAA 1170
 Db 660 CTGGAAGATATAGGAAGTATGTTGAAAGATGCGCATTTGAAAGAGGTTTCAACAG 719
 QY 1171 GTCTATGTTGGGAGGCC 1187
 Db 720 GTTATGTTGGTGTGACC 736

RESULT 11

CB893989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

EST646781 HOGA Medicago truncatula cDNA clone HOGA-29N13, mRNA

sequence.

CB893989

CB893989.1 GI:30101158

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

Medicago.

Medicago.

Medicago.

Medicago.

Medicago.

Medicago.

Medicago.

Match	Similarity	Score	DB 14;	Length	81.2;			
627;	Conservative	0;	Mismatches	176;	Indels	9;	Gaps	1;
1420	AGGATGCAGCTGGAAAATTGAACCTTCACGCTTCGGAAGGAGAGAGGATAAAGCCAGCAA	1479						
1	AGATGCAATTAGAGTTGAACTGCAATGCTCTAGAGAAAGAGAGAGCAAAAGCTAGCGAA	60						
1480	GCTCGACTTATAGAGGTGCGGAAAGAGGCTTGATGACCTGAGAGACAAGCTTCAGCCCTTC	1539						
61	GCCCGCTCTGTAGATGTACGGAGAGAACTTGATGACTTGAGAGACAAGCTTCAACCTCTG	120						
1540	ACGATGAAATACAGAAAGGAGAAAGAGAGAAATTGATGAGATTCGAAGGCTTAAACAGAA	1599						
121	AAGATGAAGTATAGTAAGAGAAAGAGAGAGATTGATGAGATTCGAAGGCTGAGCAGAA	180						
1600	AGAGAAGAGCTCAATGTTTTCTTTTCGAGGAGGAGAGAACGAGATATGACTTTGCAAGAGCT	1659						
181	CGTGAAGAGCTCCCTCTTTGCACTACAGGAGGAGAGAGCGCGTATGATCTAGCAAGAGCT	240						
1660	GCTGATCTAAGATATGCGCAATTCAGAAGTGGAACTGTGCAATTGCCAACTTGAGGA	1719						
241	GCGAECTGCGATATGGTGCAAATTGAAGAGGTGGAACTGCATTTAAAAATCTTGAAGGT	300						
1720	ACTTCT-----TCTGAAGAGAATGTGCTCTCAGAAAAAGTTGGGCGCTGACAC	1770						
301	AGCACTGATGGGAACAAGATGAGAACTTAATGTTGACGGAACAGTTGGACCGGACCA	360						
1771	ATTGCTCAGGTTGTGAGCGGTGGCAGCGGATTCAGGTGACGAGACTTGGCCAAATGAG	1830						
361	ATAGCGAGGTTGTGACCCGATGGACTGGTATACCAAGTGACTGAGCTCGCGCAAAATGAG	420						
1831	AAGGAGAGGTTGATGTCCTTCTGCTAGTATGAGTTGCAATAAGCGGTTGTGGGACAGAAATCAA	1890						

Ov. 4

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. 750
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone="BOGRL30"
/clone_lib="BOGT"
/note=vector: phos1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into phos1 using BstXI linkers"
154 a 236 c 150 g 210 t

9 GTTCTAGTCGTTATTAAGCTCATCTTCGCGCGGTCCACGGTCGACGACGACGACGACGACGATA 478

541 GTTGTGAGGATTTTGTCAAGGAGGACTAAGAAATATCCAGTCTTATTTGGTGAGCCTGGT 600
 793 GTTGTAAACAGCTGTGGTTGAAGGTTTAGCACAAGAGATTTGAAAGAGAGATGTGCC 852
 601 GTTGGAAACCTGCTGTTGTTGAAGGTTGGCTCAGAGAGATTTGAGAGGTTGTTCT 660
 853 AACAGTCTTACTGATGTCAGATTAAATTCGTTGACATGGGTGGTTAGTTGCTGGTCT 912
 661 AGCAATCTGTGATGTTAGTTAAATTCGCTCGATATGGGAGCATTTGGTTGCTGGTGG 720
 913 AAATACCGAGGAGAGTTTGAAGAAAGGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCT 972
 721 AAGTATAGGAGAGATTTGAAGAGAGGTTGAGGCTGTTTGAAGAAAGTTGAAGAGCT 780
 973 GAA 975
 ||||
 781 GAA 783

RESULT 14
 1648761
 LOCUS
 DEFINITION
 EST510380 HOGA Medicago truncatula cDNA clone phOGA-23B24 5' end,
 mRNA sequence.
 BG648761
 BG648761.1 GI:13783873
 EST.
 Medicago truncatula (barrel medic)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 818)
 Hahn,M.G., Ojano-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
 Uterback,T., Cho,J. and Frazer,C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogalacturonides of DP 6-20
 Unpublished
 Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccr.uga.edu
 G392367e TIGR sequence name: MTMCP12TK More information is
 available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gtc gat CC).
 Location/Qualifiers
 1..818
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="phOGA-23B24"
 /tissue_type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C
 with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_host="XLOLR"
 /clone_lib="HOGA"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in SOLR cells."
 239 a 151 c 230 g 198 t

FEATURES
 source

15.6%; Score 483; DB 10; Length 818;
 Query Match

Best Local Similarity 75.6%; Pred. No. 1.1e-99;
 Matches 616; Conservative 0; Mismatches 190; Indels 9; Gaps 1;
 QY 1611 CATGTTTCTTTGAGGAGGAGAGAGATATACCTTGCAGAGCTGCTGATCTAAG 1670
 Db 1 CTTCTTTGCACTACAGGAGGAGAGAGCGGTATGATCTAGCAAGAGCTGGGACCTCGG 60
 QY 1671 ATATGCGCAATTCAGAAAGTGGAAATCTGCAATTCGCCAATCTTGAAGGAATTC 1725
 Db 61 ATATGGTCAATTTGAAGAGGTGGAAATCTGCAATTTAAATCTTTGAAGGTAGCACTGATGG 120
 QY 1726 ----TCTGAAGAAGATGTGATGCTCTACAGAAACCTTTGGGCCCTGAACACATTCGTGAGGT 1781
 Db 121 GAACACGGATGAGAACTTAAATCTTGACGGAAACAGTTGGACCGGACCAATAGCCGAGGT 180
 QY 1782 TGTGAGCCGTTGGACAGGAGATTCACAGTGAACAGAGCTTGGCCAAATGAGAGAGAGGTT 1841
 Db 181 TGTAGCCGATGGATGGTATACCACTGACTAGGCTCGGCCAAATGAGAAAGCAAGGTT 240
 QY 1842 GATTGGTCTTGTGATAGTTCATAGCGGGTTTGGGACAGAGATCAAGCGGTAATATGC 1901
 Db 241 GGTGGACTTTGGTGACAGACTGCACACTAGAGTCTGGGACAAAGACCAAGCGTTAATGC 300
 QY 1902 AGTTTCTGAGGCAATTTAAGGTCAAGGGCAGGACCTTGGTAGGGCAACAAGCCCAACTGG 1961
 Db 301 TGTGTCTGAGGCTGTATTGAGATCAAGAGCTGGTTTAGGAAAGACCCCAACCAACTGG 360
 QY 1962 ATCATTTCTATTCTTGGACCAACTGGTGTGGCAAAACTCAGCTCGCCAAAGGCTCTTGC 2021
 Db 361 TTCTCTCTTATTCTTGGTCCGACTGGTGTGGTAAACTGAGCTTGCAAAAGGCTCTTGC 420
 QY 2022 TGAGCAGCTGTTGATGATGAAACCTCTTAGTTGGATGATATGTGCGGAATATATGGA 2081
 Db 421 AGAGCAGCTATTGTGATGATGAAATCAGTGTGAGAAATTCACATGTCGGAATACATGGA 480
 QY 2082 ACAACACTCTGCTCTCCCTCATTGGGGCACCACAGGAGTATGTTGGTCAAGAGGAGG 2141
 Db 481 ACAACACTCTGTATCAAGATTGATTTGGTGGCAACCACAGGGTATGTTGGACATGAGGAAGG 540
 QY 2142 TGGCAACAACTAATGAGGCTGTGAGGCGGACCTTATTGTGTCTATCTCTTTGATGAAGT 2201
 Db 541 GGGTCATTAATACAGAGAGCTGTAAAGAGAGAGCCCAATACAGTGTGATCTCTTTGATGAAGT 600
 QY 2202 GGAGAAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCCAAAGTTTGGATGATGTCGATT 2261
 Db 601 TGAGAAAGCACACACATCTGTATTCAACACTCTCTCCAAGTCTTGGATGATGGAGGTT 660
 QY 2262 GACAGCGGGCAAGGAGGAGGACAGTCGATTTTCAGGAACTCGGTGATATCATGACATCAA 2321
 Db 661 GACTGACCGACAAGGCGAGAACTGGGATTTTAGAAACACCTGTGATCATCATGACCTCTAA 720
 QY 2322 CTTTGGTCTGAACACACTCTCTTGCAGGCGCTAACTGGGAAAGTAACTGAAGTGGGCCCG 2381
 Db 721 CTTTGGTCTGAGCATCTCTTGTGAGGAGCTTTGAGGAAATGTACCATGCAAGTGTCTCG 780
 QY 2382 GGACTGTGTGATGCGGAGGAGTGAAGAAACACTTCA 2416
 Db 781 CGATCGAGTGTGATGAGGAAGTGAAGAGGCAATTTCA 815

RESULT 15
 CB893763
 LOCUS
 DEFINITION
 EST7646555 HOGA Medicago truncatula cDNA clone HOGA-28N14, mRNA
 sequence.
 CB893763
 CB893763.1 GI:30100932
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Medicago truncatula (barrel medic)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

REFERENCE
AUTHORS

1 (bases 1 to 804)
Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
Utterback, T., Cho, J. and Fraser, C.M.
ESTs from roots of Medicago truncatula treated with
oligogalacturonides of DP 6-20
Unpublished

JOURNAL
COMMENT

Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMDJ79TK

More information is available at: www.medicago.org
Seq primer: SKmod (CTA GAA CTA gty GAT CC).

FEATURES

source

1..804
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Al7"
/db_xref="taxon:3880"
/clone="HOGA-28N14"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
233 a 149 c 226 g 196 t

E COUNT
EIN

Very Match 15.5%; Score 481.2; DB 14; Length 804;
Best Local Similarity 76.1%; Pred. No. 2.8e-99;
Matches 610; Conservative 0; Mismatches 183; Indels 9; Gaps 1;

1503 GAAGAGCTCATGTTCTTTGCGAGGCGGAGACGAAGATATGACCTTGCACAGCTGCT 1662
2 GAAGAGCTCCTCTTTGCACTACAGAGGCGGAGAGGCGGTATGATCTAGCAAGCTGG 61
1663 GATCTAAGATATGCGCAATTCAAGAGTGGATCTGCAATGCCCACTTGAAGAACT 1722
62 GACCTGCGATATGGTGCATTAAGAGGTGGAACTGCAATTAAATCTTGAAGGTAGC 121
1723 TCTTCT-----GAAGAGAAATGTGATGCTCACAGAAACGTTGGGCTCAACACTT 1773
122 ACTGATGGGAACACGGATGAGAACTTAATGTGACGGAAACACTTGGACGGACCAATA 181
1774 GCTGAGGTGTCGCGCTTGACAGGATTCAGTGCAGAGACTTGGCCAAATGAGAG 1833
182 GCCGAGGTGTTAGCCGAGGACTGGTATACCACTAGTGGCTCGGCCAAATGAGAAA 241
1834 GAGAGGTTGATTTGCTCTTCTGATAGTGTGCATAGCGGTTGTTGGGACAGATCAAGCG 1893
242 GCAAGTTGGTTGGACTTGGTGACAGACTGCACACTAGAGTCGTGGGCAAGACCAAGCG 301
1894 GTAAATGACATTTCTGAGGCAATTTCTAAGTCAAGGCGAGGACTTGGTAGGGCACAACAG 1953
302 GTTAAATGCTCTTCTGAGGCTGTATTGAGATCAAGAGCTGTTTAGAGAGACCCCAAA 361
1954 CCAACTGGATCATTTCTTCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCCAG 2013
362 CCAACTGGTTCTTCTTCTTCTTGGTCCGACTGGTGTGGTAAACTGAGCTTGCAGAG 421
2014 GCTCTTGTGAGCAGCTGTTTGTATGATGAATGAATCTTCTAGTTCGGATTGATATGCGGAA 2073

Db 422 GCTCTTCAGAGCAGCTATTGATGATGAATCACTTGTGAGATTCATGTCGGA 481
QY 2074 TATATGGAACAACACTCTGCTCTCGCTCAATGGGGCCACCAAGGATATGTTGTCAC 2133
Db 482 TACATGGAAACAACACTCTGTATCAAGATTGATTGGTGACACCAAGGATATGTTGACAT 541
QY 2134 GAGGAAGGTGGACAACCTAACTGAGGCTGTGAGGAGGACCTTATTGTGTATCATCTTT 2193
Db 542 GAGGAAGGGGTCAATTAACAGAACTGTAAAGAGAAAGGCCATACAGTGTGTACTCTTT 601
QY 2194 GATGAAGTGGAGAAAGCTCATGTTGCTCTTCAACACACTCTGCTCAAGTTTGGATGAT 2253
Db 602 GATGAAGTTGAGAAAGCACACATCTGTATTCAACACTCTTCTCAAGTCTTGGATGAT 661
QY 2254 GGTCTGATTGACAGACGGGCAAGGAGGACAGTCGATTCAGGAACCTCGGTGATATCATG 2313
Db 662 GGGAGGTTGACTGACGGACAGGCGAAGCTGTGGATTTTGAAGAACACTGTGATCATG 721
QY 2314 ACATCAAACTTGGTGTGCTGAACACACTCTTGCAGGCTAACTGGGAAAGTAACTGGA 2373
Db 722 ACCTCTAACTTGGTGTGCTGAGCATCTCTTGAAGTGGACTTTTACGAAATGTACCAATGCAA 781
QY 2374 GTGGCCCGGACTGTGTGATGC 2395
Db 782 GCTGCTCGGATCGAGTGATGC 803

Search completed: February 12, 2004, 19:23:24
Job time : 5768 secs

lt	o.	Query			ID	Description
		Score	Match	Length		
1	2	663	21.4	2592	4	US-09-358-352-2604
2	2	649.4	20.9	1830121	4	US-09-527-884-1
3	3	649.4	20.9	1830121	4	US-09-643-990A-1
4	4	631.8	20.3	2628	4	US-09-107-532A-883
5	5	620.2	20.0	2574	4	US-08-887-534A-46
6	6	620.2	20.0	2574	4	US-09-527-431-46
7	7	594.8	19.2	2616	4	US-09-134-001C-1142
8	8	565.6	18.2	1230025	4	US-09-198-452A-1
9	9	546.8	17.6	2580	4	US-09-139-637A-278
10	10	546.8	17.6	2970	4	US-09-199-637A-270
11	11	545.8	17.6	42235	4	US-09-139-637A-1
12	12	545.2	17.6	2580	4	US-09-252-991A-4842
13	13	545.2	17.6	2970	4	US-09-252-991A-4897
14	14	495.4	16.0	3727	1	US-09-249-380-1
15	15	481.2	15.5	4403765	3	US-09-103-840A-2
16	16	481.2	15.5	4411529	3	US-09-103-840A-1
17	17	469.8	15.1	580073	4	US-08-545-5280-1
18	18	435	14.0	1674	4	US-09-724-623-18
19	19	376.6	12.1	2019	3	US-09-040-843-3
20	20	376.6	12.1	2019	4	US-09-621-855-3
21	21	376.6	12.1	2599	3	US-09-040-843-1
22	22	376.6	12.1	2599	4	US-09-621-855-1
23	23	370.2	11.9	2472	4	US-09-134-001C-1244
24	24	329.8	10.6	1677	4	US-09-139-637A-276
25	25	310.8	10.0	3274	4	US-09-221-017B-661
26	26	289.2	9.3	2640	4	US-09-252-991A-6033
27	27	283	9.1	2493	4	US-09-107-532A-2055

892 GGTGCGTGTAGTCTGCTGCTAAATACCGAGGAGAGTTTGAAGAAGGTTGAAATCTGTT 951
730 GGTTCATTACTTCGAGTGTAAATATCGTGTGAGTTTGAAGAGGTTTAAAGCTGTT 789
952 TTGAAAGAAGTTGAGAGCGTGAAGCGAAAGTATCTCTTATTATGATGAGATTCATTG 1011
790 TTAAGAAGATTGGCGGAAACACGAAGCGGAAATCATCTTATTCAATGACGAGTTACATACA 849
1012 GTTCTTGTGCTGGCAAACTGAAGGCTCGATGCGAGCTAAATCTGTTCAAGCCCATG 1071
850 CTCGTTGGTCTGTTAAAGTGAAGCGGATGATGCGAGGTAATATGTTTAAATCTGCG 909
1072 TTAGCTAGAGGCGAGCTTCGATGCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1131
910 TTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 969
1132 GTTGAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191
970 ATTGAAAGAGATGCGAGCTTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1029
1192 GTGCTGACACCAATTAATAGTATCTTATAGGAGCTCAAGGAGAGTATGAGGAGCATCATGCT 1251
1030 GTAGAGATACCAATTCGATTTACGCTGCTTGAAGAGAAATATGCGACTCACCCTGCG 1089
1252 GTGCGAATCCAGAGAGAGCTCTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311
1090 GTACAGATTTAGACTCAGGATTAATGCTGCGGCGAAATGCTCTCATCTGTTATTACA 1149
1312 GGTGCGCATTTACCGGATAAGCAATGATTTGTTGATGAGGCTTGTGCGAATGTGAGA 1371
1150 GACCGTCAATTACAGAGAGGCGATTTGACCTGATTTGATGAGGCGGCTTCTCGTATTAG 1209
1372 GTCCAGCTTGATAGTCAACCTGAAGAGATTTGATAACCTTGAAGAGAGGAGTGCAGCTG 1431
1210 ATGGAATTCGATTTAAGCAGAGCACTTGATAAATGATGCTGCGCTTAAATCCAGTTG 1269
1432 GAAATGAACTTCAAGCTTGAAGAGGAGAGGATAAGCAGCAAAAGCTGCACTTATA 1491
1270 AAAATGCAATTTGAAGCGGTGAAAGAGATGAAGAGCGAGGAGTAAAGCGGAGTTACT 1329
1492 GAGGTGCGGAAAGAGCTTGATCACTGAGAGAGAGCTTCAAGCTTCAAGTGAATATAC 1551
1330 CATCTTGAAATACAGATCTGATGAGTGAAGAGAGATTAACAGATCTCGAAGAGTGTGG 1389
1552 AGAAGAGAGAGAGAGAGATTTGATGAGATTTCAAGGCTTAAACAGAAAGAGAGAGCTC 1611
1390 AAGCTGAGAGAAACCTGTAGAGGCACTAAACAGCTCAGGTGGAATTTGATTAAGCA 1449
1612 ATGTTTCTTTCAGAGGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1671
1450 CGTATTGCTTTTGAAGAAAGCTCAGCGGAGAGGCGATTTGGCAGAGAGAGAGAGAGAG 1509
1672 TATGCGCAATTCAGAGAGTGAATCTGCAATTCGCAATTCGCAATTCGCAATTCGCAATTC 1731
1510 TATGCTGATTTCCAGAGTTGCAAAACCACTTGACAGAGATGAAGTGTGCTGGAAGAAAC 1569
1732 GAGAT 1788
1570 GAAGAGCAAAACTCATTCGTACAAAGATTAATGAAATGAAATGAAATGAAATGAAATGAAAT 1629
1789 CGTTGAGAGAGATTCAGTACAGAGCTTGGCCAAATGATGATGATGATGATGATGATGATGAT 1848
1630 GCTGCAACAGAGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1689
1849 CTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1908
1690 ATGAGAGAGTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1749
1909 GAGGCAATTTCAAGTCAAGGAG 1968
1750 AATGCTGTTGCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1809
1969 TTAATCTTGGACCAACTGGTGTGGCAAAACTGAGCTCGCAAGGCTTCTGCTGAGCAG 2028

1810 TTGTTTTAGACCAACAGGTTGTTGGTAAACCTGAGTTGACTAAAGCTTTAGCTAACTTC 1869
2029 CTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
1870 TTGTTGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
2089 TCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2148
1930 TCTGCTGAGCGCTTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1989
2149 CTAAGTGAAGCTGTGAGAGAGCGGACCTTATTTGTTGCTCATCTCTTTGATGATGATGATGATGAT 2208
1990 TTAACCGAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2049
2209 GCTCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2268
2050 GCGCATCCAGATGCTTCAATATCTTGTCTACAAAGTGTAGACGATGAGCGCTTAAACCGAC 2109
2269 GGGCAAGCGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328
2110 TCACAGGCT 2169
2329 GCTGAAACACCTCTTGCAGGCTTAACTGGGAAAGTAACTGGAAGTGGCCCGGACTGT 2388
2170 TCACAGATGCT 2226
2389 GTGATGCGGAGGAGTGAAGAAACCTTCAGACAGAGCTTTTGAACAGGCTTGAACAGAT 2448
2227 GTAATGAATCGGTAACTGAGCATTTCCGTCAGAAATTTATTAACCGGATTTGATGATGATGATGAT 2286
2449 GTGCTGTTCCAGCCCTTTTCAATGATGAGCTTGAAGAGAGTGAAGTGGCTTCAATGAA 2508
2287 GTGATTTTCACTTCACTTCAAAAGAGAGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2346
2509 GAGTGTGCT 2568
2347 CGCTTACGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2406
2569 GACTATATCTTGGCAGAGATTTATGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2628
2407 GACTTATGAT 2466
2629 ATGAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2688
2467 ATTCAACAGAGTGAAGAAATACACTAGCTCAAAATCTTGTGAGGCTGCTGCTGCTGCTGCTGCTGCT 2526
2689 AACTCCACTGTTTACATAGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2731
2527 GGTGATACCATTTTAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2569

RESULT 2

US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22

Query Match	20.9%; Score 649.4; DB 4; Length 1830121;
Best Local Similarity	56.6%; Pred. No. 3.4e-176;
Matches 1243; Conservative	0; Mismatches 946; Indels 6; Gaps 2;
426	TCCTATTAAAGTCATTCTGTCGTCTCAAGCTGCTCAGAAGTCACGAGGTGATCTCATTT 485
906639	TATTAACTCTAATTAAATTTATGCGATAAATTCGACACGCAAAACCAAGATAAATTTATTTTC 906698
486	GGCTGTTTGACACAGTTGATTATGGTCTTCTTTGAAGATTCTCAATACAGGGATTTCTTGAA 545
906699	GAGCGAATTGTTTGTTCGACCTTTAGAAAGAACGAGGAACGATCAGCGATATTTTGAA 906758
546	GGAAGTCGGTTGACGACGCGGAGGGTAAAGTCTGAGTTGTGAAGCTTCGTGGGAAAGA 605
906759	AAAGTGGCGTTCGCAAAAAGAAACAAATTTTCGAAGCTATTTCAGCATTTAGAGSGGGACA 906818
606	AGGGAAGAAAGTTTCAGAGTCTTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGG 665
906819	AAACGTGAACGATCAAAATTCAGAAAGAAAGCAGACAAAGCGCTTGAAAAATATACGATTGA 906878
666	AAGAGATTTTGTTGAGCAAGCAGGGAAGCTTGATCCTGTGATTGTTGTCGTGATGAGAGAT 725
906879	TTTAAACGCTCGTGCAGAAAGTGCACATTTGATCTGTAAATGGCGGTGATGAAGAAAT 906938
726	TAGAAGAGTCGTGAGGATTCCTTTCGAGGAGAAACGAAGAACATCTCTGTCTATTTCGAGA 785
906939	TCGTCAGGCCATTCAAAGTATTCAACGTCGTACCAAAAATAACCTGTGTTAAATGGTGA 906998
786	GCACAGAGTTGGTFAAAACAGCTGTGGTTTCAAGGTTTACGACAAAGGATTTGGAAGGAGA 845
906999	ACCAAGTGTAGGGAAGAACCGCGAATTTGCAGAGGCTTGGCAGACGCCATCGTAAACCGCGA 907058
846	TGTGCCCAAACAGTCTTACTGATGTGAGATTAAATTTCTGTTGGACATCGGTGCTTAGTTGC 905
907059	AGTGCAGAAAGGTTTGAAATAAACAAGTGTCTTTCAATTAGATATGGGGCGTTGATTCG 907118
906	TGCTGCTAAATACCGAGGAGAGTTTGAAGAAAGTTGAAATCTGTTTTGAAGAAAGTTGA 965
907119	TGCTGCGAAATATCTGTGTGTGAATTTGAAGACGTTTAAAGCAGTACTCAATGAACTTTC 907178
966	GAACGCTGAAGGCAAAAGTCATTTCTTTTATTCATGAGATTCAATTGCTTTCTGTGCTGG 1025
907179	GAAGAGAAAGTCCGGTTATCCTCTTTATTCGACAAATTCATCTATGTCGCGCGGG 907238
1026	CAAACTGAAGGTCGATCGATCGACTAAATCTGTTCAAGCCCATGTTAGCTAGAGGCA 1085
907239	TAAACCGATGTCGATGATGCGGGTAAATTTGTTAAACCAAGTTTGGCACGAGCGCA 907298
1086	GCTTCGATGCATTTGGTCTTACCAACGCTTGAAGAAATACAGGAAATATGTTGAGAAAGATGC 1145

1326 GGATTAAGCAATTGATTTGGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTGTGATAG 1385

907539 AGATTAAGCCATTTGATTTGATGATGAAGCAGCGCTAGCATTTGATTCGTAATCGAATAAGATTTC 907598

1386 TCAACCTTGAAGAGATTGATAAACCCTTGAAGAGAGAGATGCAGCTCGAAATTTGAATCTTCA 1445

907599 TAAACCTGAACCGCTTGATTCGTTGNAACGTGCTATTATCCATTTAAATTTGGAACTCA 907658

1446 CGCCTTGGAAAGGGAGAGGATTAAGCCAGCAAAAGCTTCGACTTTATAGAGGTGCGGAAGA 1505

907659 AGCGTTACAAAAGAAAGAACACCAAGCAAGCTCGCAACGTTTGAATACTGTAGAGAAAGA 907718

1506 GCTTGTGATGACCTTGAGGACACAGCTTCAGCCTCTCACGATGAAATACAGAAAAGGAGAAAGA 1565

907719 ATTGGCTGAAAAGAACGCTGAATACGCCGAACTTTGAAGAAGTATGAAATCTGAAAAGAC 907778

1566 GAGAATTGATGAGATTCGAAGGCTTTAAACAGAAAAGAGAGAGCTCATGTTTCTTTTGCA 1625

907779 AACGCTTCTGGCTCTCAACATATTAACCAAGAGTTAGATACTGCAAAAACCGAATCTAGA 907838

1626 GGAGGACAGAAAGATATGACCTTGCAGAGAGCTGCTGATCTAAGATATGCGCGCAATTTCA 1685

907839 ACAAGCTTCGTGCGCGGGTGATTTAGCGAAATGTCTGAAGTTTGCATATATGGCCGCAATCC 907898

1686 AGAAGTGAATCTGCAATTCGCCAACTTTGAAGGAACTTCTCTGAAAGAAATGTGATGCT 1745

907899 TGATCTTGAAGACCACTTTGAGCAAGCTGAAACCGAGGAGGAAAGAAATGACGCTTTT 907958

1746 CACAGAAAACGTTGGGCGCTGAAACATTTGCTGAGGTTGTGAGCGGTGAGCAAGGAAATCC 1805

907959 ACGCTATCGCGTCAACAGATGAAGAAATCGCAGAAGTGCTTTCTAAAGCCACAGGCAATCC 908018

1806 AGTGACGAGACTTTGGCCAAAATCAGAAAGAGAGGTTGATTTGGTCTTCTGATAGGTTGCA 1865

908019 TGATCAAAAATGATGAAGCGCAGAAAGAAAACCTCTTCGCTATGGAAGATGAATCACTCA 908078

1866 TAAGCGGTTTGTGGGACAGAAATCAAGCGGTAAATGCAAGTTTCTGAGGCAATTTCTAAGGTC 1925

908079 TAAACGAGTGAATTTGGTCAAGAAAGCGGTTGATGCGGTAGCAAAACGCGATTTCTGCTGAG 908138

1926 AAGGCGAGACTTTGTTAGGCAACACAGCCAACTGATCATTTCTTATTCCTTTGGACCAAC 1985

908139 TCGTGCAAGTCTTTTCCGATCCTAATCGCCCAATTGGTTCTTTCTGCTTCTTAGGCGCAAC 908198

1986 TGGTGTGGCAAACTGAGCTCGCAAGGCTCTGTCTGAGCAGCTCTTTGATGATGATGAAAA 2045

908199 AGGTGTTGGGAAACAGAGCTTTGCAAAACCTTTGGCTAAATCTTGTTGATAGTGAAGA 908258

2046 CCTCTTAGTTGCGAATTGATATGTCGGAATATATGGAACAACACTCTGTCTCTCGCCTCAT 2105

908259 TGGATGTCGTCGATTTGATATGTACAGATTTATGAAAAACACAGTGTGTTCTCGTTTAGT 908318

2106 TGGGGCACCAACAGGCTATCTGGTACAGAGGAAGGTGACAACTAACTGAGGCTGTGAG 2165

908319 TGGTGGCCTCCAGGCTATGTCGGCTATGAAAGAGCGGTTATTTAACTTGAAGCTGTTGCG 908378

2166 GAGCGCACTTATTTGTTGCATATCTTTGATGAAGTGGAGAGGCTCATGTTGCTGTCTTT 2225

908379 TCGTCGTCCATATTCAGTGTATCTTATAGATGAAGTTGAAAAAGCACACGCAATGTATT 908438

2226 CAACACTCTGCTCCAGTTTGGATGATGCTGATTTGACAGCGGACAGGACAGGACAGT 2285

908439 CAATATCTTATTAACAGTGTGATGATGCTGTTAACTGATGGTCAAGGCTGCTACTGT 908498

2286 CGATTTTCAGAACTCGGTGATATCATGACATCAAACTTTGGTGTCTGAAACACTCTCTTCG 2345

908499 GGACTTCGGTAACACTGTGGTTATATGACCTCTAATCTGGGTTCTGATTTAATCCAAG- 908557

2346 AGGCTAACTGGGAAAGTAACAATGGAAGTGGCCCGGACTGTGTGATCGCGGAGGTGAG 2405

908558 -GTAAATAAAGCAAGAAAGCTATAGCGAAA----TGAAGGCTTATGATGATCTCAGTGGTAA 908612

2406 GAAACACTTCAGACCAAGAGCTTTGAAACAGGCTTTGACGAGATTTGCTGTGTCGACCCCT 2465

Db	908613	CCAA	CATTTC	CGCC	CAGAA	TTCAT	CAAC	CGTAT	TGAC	GAAC	CGTG	TATTC	CATC	CACT	908672	
Qy	2466	TTCA	CATG	ACCA	GAGT	AGTAG	CTCG	CTTCA	AAAT	GA	AGAC	GTG	CTCT	CGGCT	2525	
Db	908673	TGTA	AGAA	AAAT	ATCC	GTG	CAAT	TGCA	AGCT	TAG	CAAA	CGTAT	908732			
Qy	2526	TGCT	GAAG	AGAG	GTG	CTTT	GGC	AGTCA	CTG	ATG	CTG	CTTT	TGGA	CTATAT	CTTTGGGAGA	2585
Db	908733	GGAA	CTCG	TGAT	ACGA	TAAT	TGG	TGTT	TAC	CGAT	GTCT	TATT	TAGAC	TCTT	CATTGGCGAAGT	908792
Qy	2586	GAGT	TATG	ACCG	GTG	TATG	TGCT	AG	CCCT	TATAA	2620					
Db	908793	GGGA	TACG	ACCA	CCAA	TTAT	TG	TGTC	AG	CGCT	CAAT	TGA	908827			

RESULT 4

US-09-107-532A-883
Sequence 883, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 883:
SEQUENCE CHARACTERISTICS:
LENGTH: 2628 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...2628
SEQUENCE DESCRIPTION: SEQ ID NO: 883:
US-09-107-532A-883

Query Match 20.3%; Score 631.8; DB 4; Length 2628;
Best Local Similarity 57.3%; Pred. No. 1.1e-172;
Matches 1183; Conservative 0; Mismatches 877; Indels 6;

583 AACGCGGAAGCTTGATCTGTGTATGTGCTGTGATGAGGAGATTAGACAGATCGTGAGGA 743

542 AAACGCGGAAAAATGGATCCAATCATCGGTCGTGTATGAAGAAATCCCGTGATGTCATCCGTA 601

743 TTCTTTTCGAGGAGAACGAGAACAAATCTGTGCTTATTTGGAGAGCCAGAGTTGCTAAAA 802

602 TTCTTTTCAGAAAAACAATAATTAATCTGTCTGTATAGGAGAACCCGGTGTAGGGNAAA 661

803 CAGCTGTGTTGAAGGTTTAGCACAAAGGATTTGAAAGAGAGATGTGCCCAACAGCTTTA 862

662 CGGCAATCGTTGAAGGGCTGCTCAACGTAATCGTACGCAAGATGTGCTCGAAATTTAA 721

863 CTGATGTGAGATTAATTTCTGTTGGACATGGGTGGTTAGTTGCTGTGCTATAACCGAG 922

722 AAGATAAAACGATTTTTTTCATTAGATATGGGTGCAITGTATCGCTGAGCAAAATTTCCGTG 781

923 GAGAGTTTCAAGAAAGGTTGAAATCTGTTTTGAAAGAGGTTGAGGACGCTGAAGCAAG 982

782 GTGAATTTGAAGAAAGACTAAAGCAGTCTTTGAAGAGGTCAAAAAAGTGATGACGTA 841

983 TGATTTCTTTATTGATGAGATTCATTTGGTTTCTGTGCTGGCAAACTGAAGGGTCGA 1042

842 TCTCTTTTCATCGATGAATCCACAATATCGTAGGGCAGGAAAAACAGAAAGAGTA 901

1043 TGGATGCAGCTAATCTGTTCAAGCCATGTTAGCTGAGGCGAGCTTGATGCAATGGTG 1102

902 TGGATGCGGAAATCTGTTGAACCAATGCTGCTCGCGTGAAITACATTTGATCGGTG 961

1103 CTACAACGCTTCAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGT 1162

962 CCACAATTTAGATGAGTACCGTCAATATATGGAAGAAAGCAAGCGTTGGAACGTGCGT 1021

1163 TCCAAACAAGTCTATGTTGCGGAGCCAAAGTGTGCTGACACCAATTAGTATCTCTTAGAGAC 1222

1022 TTCAAAAGGTATTAGTCAAGSAAACCAACAGTGGAGATACGATCAGTATTTCTGCGTGAT 1081

1223 TCAAGGAGAGATATGAGGGAATCATGTTGTGCGAATCCAAAGACAGAGCTCTTATAAATG 1282

1082 TAAAGAAAGCTTTTGAGATCCACATGGTGTAAACATCCATGCAACGCAATTAGTAGCAG 1141

1283 CTGCTCAGCTGTCTGCTCGTTTACATACTGGTCGGCATTTACCGGATTAAGCAATTGAT 1342

1142 CAGCTACTTTATCTGACCGGTTATTAATCTGATCGGTTCTTACCGGACAAAGCAATTCGAT 1201

1343 TGGTTGATGAGGCTTTGTGCGAATGTGAGAGTCCAGCTCTGATAGTCAACTGAAGAGATTG 1402

1202 TGATCGATGAAGCCAGTGCACAGATTCGGGTAGAAATGAACTCTATGTCACCGGAATTAG 1261

1403 ATAACTTTGAAGAGAGAGATGACGTGAAATGAACTTCAACGCTTTGAAAGGGAGA 1462

1262 ATCAAGTTTACTAGACGTTTGTATGCAGTTGGAATTCGAAGAAGTCACTGAAAAAGAA 1321

1463 AGGATAAAGCCAGCAAGCTCGATTAATAGAGTGGGAAGAGCTTGATCACTCTGAGAG 1522

1322 CAGATGATGCAAGTAAAAAACGTTGAAAAATTTTGAAGAANAATTAGCTGAATCTCGTG 1381

1523 ACAAGCTTTCAGCCTCTCACGATGAAATACAGAAGGAGAAAGAGAAATTCATGAGATT 1582

1382 AAGAAGCAAAATGCCATGNAAGTGCATGTTGGGAAACAGAAAAGAGAGTCAATTCGTGTT 1441

1583 GAAGCTTAAACAGAAAAGAGAGAGCTCATGTTATGCGGCAATTCAGAAGTGAATCTGCAA 1642

1442 CAGCAAAAACGTGCTGAAATCGATAAAGCAAAACACGAATTTAGAAATGCAGAAAAAAT 1501

1643 ATGACCTTTGCAAGAGCTGCTGATTAAGATATGCGGCAATTCAGAAGTGAATCTGCAA 1702

1502 ATGATTTAGAACCGCGCTGCTGTTATTCGCTCATGTTGACGATACCTCAATTTAGAAAAAG 1561

1703 TTGCCCAACTTTGAAGGAACCTTCTCTGAA --- GAGAAATGTGATGCTCACAGAAAAACCTTG 1759

1562 TGAAGGATTAGAGCTTAAGCAAAAGACAGCGAGATTAAGTGTTCAGAGTCCGTTA 1621

1760 GGCCTGAACAATTTGTCGAGGTTGTGAGCCGTTTGAACAGGGAATTCAGTGAACAGACTTG 1819

Db	1622	CCGAAAATGAGATTGCTTCAAGTCGTTGGCAGATTAAACGGGTATCCCGTAAACGAAACTTGG	1681
QY	1820	GCCAAAATGAGAAAGCAGGTTGATTGGTCTTGCCTGATAGTTGCAATAAGCGGTTGTCG	1879
Db	1682	TAGAAGGTGACGAGAAAACTAATTAAATTGAATGAACACTTCAATAACGAGTGAATTG	1741
QY	1880	GACAGAAATCAAGCGGTAAATTCGATGTTCTGAGCAATTTCAAGGTCAAGGCGCAGGACTTGG	1939
Db	1742	GTCGAAGATGAAGCAGTTGATCGCGTCAGTCGACGCGTAAATTCGTTCAACGGGCTGGATTGC	1801
QY	1940	GTAGGGCACACAGCCAACTGGATCATTTCTATTCTTGGACCAACTGGTGTGGCAAAA	1999
Db	1802	AAGACCCAAATCGTCTCTAGGTTCTTTCTTCTCTAGGACCAACTGGTGTGGCAAAA	1861
QY	2000	CTGAGCTCGCAAGGCTCTTGCTGAGCAGCTGTTGATGATGAAAACTCTTAGTTTCGGA	2059
Db	1862	CCGAATTAGCGAAAGCGTTAGCTGAGAACTTGTGTTGACTCAGNAGATCATATGTTACGGA	1921
QY	2060	TTGATATGTCGGAATATATGGAACAACACTCTGTCTCTCTCCCTCAFTTGGGGCACCAACG	2119
Db	1922	TCGATATGAGTGAATATATCGAAAAACATGCTGTGCTCGTCTAGTCGGAGCTCTCCAG	1981
QY	2120	GGTATGTTGTTCAGGAGGTTGGCAACTACTCAGGCTGTGAGGAGCGCACTTATT	2179
Db	1982	GCTATGTAGTTTACGAAGAAGTGGTCAATTAACGGAAGCTGTCAAGCAAAATCTCTATA	2041
QY	2180	GTGTCAATCTTTTGATGAAGTGAGAAGGCTCATGTTGCTCTCTTCAACACTCTGCTCC	2239
Db	2042	CAATCGTCTTTGGATGAATCGAAAAAGCGCATCCCGATGCTTCAATACTTGTGTAC	2101
QY	2240	AAGTTTGGATGATGTTGATTCACAGACGGGCAAGGCAGGACAGTCGATTTTCAGGA	2299
Db	2102	AAGTCTTAGATGACGACGCTTGACTGATTTCAAAAGGCGGAGTAGTAGATTTCAAAAATA	2161
QY	2300	CGGTGATAATCATGACATCAAACTTTGGTGTCTGAACACCTCTCTCAGGCGCTAACTGGGA	2359
Db	2162	CGGTTCTGATCATGACCGAATATCGGTTACAGTTACTGCTGGAAGCGTGA	2221
QY	2360	AAGTAACAAT- --GGAATGGCCCCGGAATGTGTGATCGGAGGTGAGGAAACATTTCA	2416
Db	2222	ATGMAACGATTCACAGAAGCAGTGGCTGAAACAAGTCAACACGCTATTACGTGGAACATTTCA	2281
QY	2417	GACACAGCTCTTTGAACAGGCTTGACGAGATTTGGTGTTCGACCCCTTTCCATGACC	2476
Db	2282	AACCAAAATCTTTGAACCGAAATCGACGACACAGATTTTATTCACCAATTGAGCCCTAGACA	2341
QY	2477	AGTTTGAGGAAAGTAGCTCGGCTTCAAAATGAAAGAGTTCGTCTCGGCTTGTCTGAAAGAG	2536
Db	2342	ACGTGAAAGGAATCGTTGACAAATGGTGGCCAAATGGGCAACAGCTCTGGAACATCAAG	2401
QY	2537	GAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGAGGTTATGACC	2596
Db	2402	AAATCTGCTGCAATCAGTGAATGAAGCCAAACCGTGAATTCGTGAAATGCTTTATGAAC	2461
QY	2597	CGGTGTATGGTGAAGCCCTATAAGGAGATGGAATGGAGAGAAAGGTGTGTGACAGACTGT	2656
Db	2462	CAGCATATGGTGCAGACCAATTTGAACAGATTTATCACTAAGGAAGTAGAAACACCACTAG	2521
QY	2657	CAAAAGATGTTGTGGTGAAGGAAATCGATGAAATCTCCACTGTTTACATAGATCGAGCG	2716
Db	2522	CTAAAGAAATCGTTGCAAGACATGTTATGCTCTAATCAAAAGTAACGATTTACTTACTTG	2581
QY	2717	CTGGTGATCTTGTACCGGGTAGAA	2742
Db	2582	ATGGGCAACTTCATTTTCAAAAACAGAA	2607

RESULT 5
US-08-887-534A-46
; Sequence 46, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:

APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Garstein, Murray & Borum
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/887,534A
 FILING DATE:
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 28341/33996
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 474-6300
 TELEFAX: (312) 474-0448
 TELEX: (312) 474-6600
 INFORMATION FOR SEQ ID NO: 46:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2574 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2571
 88-887-534A-46

Query Match	20.0%; Score 620.2; DB 4;	Length 2574;
Best Local Similarity	54.7%; Pred. No. 2.4e-169;	
Matches 1278; Conservative	0; Mismatches 1053; Indels	6; Gaps 2;
352	AGAGTGATCAATCAAGCTTGAAGAAGCTTCCTTCACAATCTCCTCAACCTGATGATATT	411
184	ACAGATATCAATCAGGCAATTAAATTCGTTTACCGCAGGTTGAAGGTACTGGTGGTGAATGC	243
412	CCAGCGAGTTCCTAGTCTTTATTAAGGTCATTTCGTCTGCTCAAGCTGCTCAAGACTCAGGA	471
244	CAGCCATCACAGGATCTGGTGC CGCTTCTTAATCTTTCGCAACAGCTGCGCGCAAAAAGCT	303
472	GGTGATATCTATTTCGGCTGTTGACCAAGTTGATTATGATGGGTCTCTCTTGAAGATCTCAATC	531
304	GGTGATAACTTATCTCGTCAGCAACTGTTGCTTCTGCGCGCACTTGAGTCTCGCGGCACC	363
532	AGG---GATTGTTGAAACGAAGTCGGTGTAGCGACGGCGAGGTAAGTCTGAGGTTGAG	588
364	GTGGCCGACATCTCTGAAGACAGCAGCGGGCGCACACCGCCAACTATTACTCAAGCGATTGAA	423
589	AAGCTTCGTGCGGAAAGGAGGAAGTTGAGAGTGCTTCAGGGGACACAAATTTTCAA	648
424	CAATCGTGGAGGTGAAACGTTGAACGATCAAGGTGCTGAAGACCAACGTCAGGCTTTC	483
549	GCTTTAAAGACTTATGGAAGAGATTTGGTTGACGAAGCAGGGAAGCTTGAATCTGTGATT	708
484	AAAAAATATACCATCGACCTTACCGAAACGAGCCGAAACAGGCAAACTCGATCCGCTGATT	543
709	GGTCGTGATCAGGAGATTAGAAGTCGTGAGGATCTCTTCGAGGAGAAACGAAGACAAAT	768
544	GGTCGTGATCAGGAATTCGCCCTGCATTCAGGTGCTGCAACCGTCGTACTTAAATAAAC	603
769	CTGTGCTCTTATCGAGAGCCAGGATTTGGTAAACACAGCTGTGGTTTGAAGGTTTAGACACAA	828

Db	604	CCGGTACTGATTGGTGGAAACCCGGCGGTGGTAAAACTGCATCGTTTGAAGGCTGCGCGCAG	663
QY	829	AGGATTGTGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTGTTGCAC	888
Db	664	CGTATTATCAACGGCGAAGTCCCGGAGGGTTGAAGCCCGCGGTACTCGCGCTGGAT	723
QY	889	ATGGGTGCGTTAGTTGCTGGTGCTAAATACCAGGAGAGAGTTTGAAGAAAGGTTGAAATCT	948
Db	724	ATGGCGCGCGCTGGTGGCTGGCGCGAAATATCGCGGTGAGTTTGAAGACGTTTAAAGGC	783
QY	949	GTTTGTGAAGAAGTTGAGGACGCTGAAGGCGAAAGTGATCTCTTTTATTGATGAGATTCAT	1008
Db	784	GTGCTTTAACGATCTTGGCAAAACRGGAAAGGCAACGTCACTCTATTATTCGACGAATTAAT	843
QY	1009	TTGGTTCTTGTGTGTGGCAAAATCTGAAGGGTTCGATGATGACGACTAATCTGTTCAAGCCC	1068
Db	844	ACCATGTTGCGCGGGTTAAAGCCGATGGCGCAATGACGCCGGAATACTGCTGAACACG	903
QY	1069	ATGTTAGCTAGAGGCGAGCTTCGATGCATTGGTGCTACAACGCTTGAAGAATA CAGGAAA	1128
Db	904	CGCGTGGCGCTGTGTAATTCGACTGCTAGGTGCGCAGCGCTTGACGAATATCGCCAG	963
QY	1129	TATCTTCGAAAGATGCTGCTTTGACGAGAGGTTTCCAAACAGTCTATGTTGCGGAGCCA	1188
Db	964	TACATTTGAAGAAAGATGCTGCGCTGGAAACGTCGTTTCCAGAAAGTGTGTTGTCGCGACCT	1023
QY	1189	AGTGTGCTCTGACACCAATTAGTATCTCTTAGAGGACTCAAGGAGAGATGATGAGGACATCAT	1248
Db	1024	TCGTTCGAAGATACCATTCGATTCGCTGGCGCTGAAAGAACGTTACGAAATTCGACCCAC	1083
QY	1249	GGTGTGCGAATCCAAGACAGAGCTCTTTATAAATGCTGCTCAGCTGTCTGCTCGTTACATA	1308
Db	1084	CATGTGCAAAATTACTGACCGCGCAATTTGTTGACGCGGAGGTTGTCTCATCGCTACATT	1143
QY	1309	ACTGTCGGCATTTACCGATAAGCAATTGATTTGGTTGATGAGGCTTGTGCGAATGTG	1368
Db	1144	GCTGACCGTCACTGTCGGGATAAAGCCATCGACTGATCGATGAACGAGCATCCAGCATT	1203
QY	1369	AGAAGTCAGCTTGATAGTCAACCTGAAGAGATGTATAACCTTGAAGGAAGAGGATGCGAG	1428
Db	1204	CGTATGCGATTGACTCAMAACAGAGAAGTACTCGACCGACTCGATCGTGTATCATCCAG	1263
QY	1429	CTGGAATTTGAATCTCAGCGCTTTGGAAGGGAGAGGATAAAGCCAGCAAGCTCGACTT	1488
Db	1264	CTCAACTGGAACAAACAGCGCTTAATGAAAGAGTCTGTATGACGCCAGTAAATAAAACGTCTG	1323
QY	1489	ATAGAGTGGGAAAGAGCTTGATGACCTGAGAGACAAGCTTCAGCCTCTCACAGATGAAA	1548
Db	1324	GATATGCTCAACGAAGAACTGAGCGCAAAAGAAACGTCAGTACTCCGAGTTAGAGAAGAG	1383
QY	1549	TACAGAAAGGAGAAAGAGAGAAATTGATGAGATTCGAGAGCTTAAACAGAAAGAGAGAG	1608
Db	1384	TGGAAGCAGAGAAAGCATCGCTTCTGTATCGAGACCAATTAAGCGGAACTGGAACAG	1443
QY	1609	CTCATGTTTCTTTGAGGAGGCAGAACGAAGATATGACCTTTGCAAGAGCTGCTGATCTA	1668
Db	1444	GCGAAAATCGCTATTGAAACAGGCTCGCCGTGTGGGGAGCTTCGCGCGGATGTCTGAATG	1503
QY	1669	AGATATGGCGAATTCAGAGAGTGGAAATCTGCAATTTGCCCAACTTTGAAGGAACTTCTTCT	1728
Db	1504	CAATACGGCAAAATCCGGAACTGGAAAGAACCACTGGAAGCGCAACGACGCTCGAAGGC	1563
QY	1729	GAGAGAAATGTGATCTCACAGAAACGTTGGCCCTGAAACATTTGACGAGGAGGTTGTGAGC	1788
Db	1564	AAACATATCGGCTCTGTTCGGTAAATAAGATGACCGGACCGCGGAAATTTGCTGAATGCTGGCG	1623
QY	1789	CGTTGGACAGGGATTCAGTACAGAGACTTGGCCAAATATGAGAGGAGGAGGTTGATTCGT	1848
Db	1624	CGTTGGACGGGATTCGGGTTCTCGCATGATGGAAGCGGCGGCAAAACTGCTGCGT	1683
QY	1849	CTTGTGTAAGGTTGATGAAGCGGTTGTGGGACAGAATCAAGCGGTAAATGCAAGTTTCT	1908

1684 ATGGAGCAAGAACTGCAACATCGCGTAATGTCAGAAAGAAAGCGGTTGATCGGTATCT 1743
 1909 GAGGCAATTTCTAAGGTCAAGGGCAGGACTTGTAGGCAACACAGCAACCACTGATCATTC 1968
 1744 AAGCTATTTCGTAGCGGTGGGGCTGGCGGATCCAAATCGCCGATGTTGTTTCATTC 1803
 1969 TTATTCCTTGGACCAACTGTTGTTGGCAAACTGAGCTGCGCAAGGCTCTTCTGAGCAG 2028
 1804 CTGTTCTCGGCCCAACTGTTGTTGGGGAAGAAAGAGCTTTGTAAGCGCTGCGCAACTTT 1863
 2029 CTGTTTGTATGATGAAGAACTCTTAGTTCGATTTGATATGTCGATATATGTAACAAAC 2088
 1864 ATGTTTGTATGAGCAAGAGCGATGTTGTCGATATGTCGATATGTCGATATGTCGAT 1923
 2089 TCTGTTCTTCGCTCATTTGGGCAACACAGAGGATGTTGTTGTCAGAGAAAGTTGACAA 2148
 1924 TCGGTGTCGTTTGGTTGGTGGCTCGGATATGTCGATATGTCGATATGTCGATATGTC 1983
 2149 CTAACCTGAGGCTGAGGAGGCACTTATTTGTCGATATGTCGATATGTCGATATGTCG 2208
 1984 CTGACCGAAGCGGTGCTGCTCGCTGATTTCCGTCATCTGCTGATGAAGTGAAGAA 2043
 2209 GCTCATGTTGCTGCTCTCAACACTGCTGCTCAAGTTTGGATGTCGATGTCGATGTCG 2268
 2044 GCGCATCGGATGCTTCAACATCTGTTGAGGCTGTCGATGTCGATGTCGATGTCG 2103
 2269 GGGCAAGCAGGACAGTCCGATTTCAAGAACTCGGTGATATGTCGATGTCGATGTCG 2328
 2104 GGGCAAGGAGAGAGCGTTCGACTTCCGTAATACGCTGCTGATGTCGATGTCGATGTC 2163
 2329 GCTGAACACCTCTTTCAGGCTTAACCTGGAAGTAACTGGAAGTGGCCCGGACTGT 2388
 2164 TCCGATCTGATTCAGAAAGCTTCGTTGAACTGATTAATGTCGATGTCGATGTCG 2220
 2389 GTGATGCGGAGGTCAGGAAACACTTTCAGACAGAGCTCTTGAACAGCTTTCAGAGAT 2448
 2221 GTGCTCGGTGTTGAGGCACTTTCGCTCGGATTAATTCATTAACGATGTCGATGTC 2280
 2449 GTGGTTCGACCCCTTTCATGACAGCAGTTCAGTTCAGGAAAGTTCGCTGCTCAATG 2508
 2281 GTGGTCTTCCATCCGCTGGTGAACAGCAGTTCGCTGATGTCGATGTCGATGTCG 2340
 2509 GACGTTGCTGCTCGGCTTTCGAAAGAGGCTGTTGTCGAGTTCATGCTGCTGCTG 2568
 2341 GCTGTTCAAAACGCTTCGAAAGAGGCTGTTGTCGAGTTCATGAAATTCGATGTCG 2400
 2569 GACTATATCTTGGCAGAGATTCACCCGCTGATGTCGATGTCGAGTTCATGAAATTC 2628
 2401 AAACCTGTCAGCAGAGAGGTTACGATCCGCTGATGTCGATGTCGATGTCGATGTC 2460
 2629 ATGGAGAAAGGTTGGTGAAGAACTGTCAGAAAGTTCGATGTCGATGTCGATGTCG 2685
 2461 ATTCAGCAGCAGATCGAAACCCGCTGTCGACAGCAAAATGTCGATGTCGATGTCG 2517

RESULT 6

US-09-527-431-46
 Sequence 46, Application US/09527431
 Patent No. 6485899

GENERAL INFORMATION:
 APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA: US/09/527,431
 APPLICATION NUMBER: US/09/527,431
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA: US/08/887,534
 APPLICATION NUMBER: US/08/887,534
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 28341/33996
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 474-6300
 TELEFAX: (312) 474-0448
 TELEX: (312) 474-6600
 INFORMATION FOR SEQ ID NO: 46:
 LENGTH: 2574 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2571
 US-09-527-431-46

Query Match 20.0%; Score 620.2; DB 4; Length 2574;
 Best Local Similarity 54.7%; Pred. No. 2,4e-169;
 Matches 1278; Conservative 0; Mismatches 1053; Indels 6; Gaps 2;
 QY 352 AGAGTATCATCAATCAAGCCTTGAAGAGCTTCTTCAATCTCTCCACCTGATGATTT 411
 DB 184 ACAGATATCAATCAGCAATTAATCGTTTACCGCAGGTTGAAGTACTGTTGATGTC 243
 QY 412 CCAGCGAGTCTTAGTCTTATTAAGGTCATTCTGCTGCTCAAGCTCTCAGAGTCTC 471
 DB 244 CAGCCATCAGAGTCTGTTGCGGTTCTTAACTTTGCGCAACGTTGGCGCAAAAAGT 303
 QY 472 GTGATATCATTTGGCTGTTTGAACAGTTCATTTGAGTTCCTTTGAAAGATTTCTCAA 531
 DB 304 GGTGATAAATTTATCTGCTCAGACTGTTCTGTTCTGGCGCACTTGAAGTCTCGCG 363
 QY 532 AGG---GATTGTTGAACGAACTCGGTGAGCAGCGGAGGTAAGTCTGAGTTCGAG 588
 DB 364 GTGGCGCAGCATCTGAAAGCAGCGGCGGCGACCCGCGCAACATTTACTCAAGCG 423
 QY 589 AAGCTTCGTGGGAAAGAGGGAAGAAAGTTGAGAGTTCAGGGGACACAAATTTTCAA 648
 DB 424 CAAATCGGTGAGGTTGAAGCTGTAACGATCAGGTGCTGAGAGCAACAGTTCAGC 483
 QY 649 GCTTTAAGACTTATGGAAGAGTTGTTGAGCAAGCAGGGAAGCTTCCTGTTGATT 708
 DB 484 AAAAAATATACCATCGACTTACCAGAAAGCGGCGGAAAGCGGCAACTCGATCCG 543
 QY 709 GGTCTGATGAGGAGATTAAGAGTCTGAGGATTTCTTCGAGAGAGCAACAGACAA 768
 DB 544 GGTCTGATGAAGAAATTCGCGTACCATTCAGGTGCTGCAACCGTCTGCTGCTGAT 603
 QY 769 CCTGTGCTTATTTGAGAGCAGGAGTTCGTTAAACAGCTGTTGAGGTTTGAAGCA 828
 DB 604 CCGTACTGATTGTTGAACCCCGGCTCGTAAACTGCACTGTTGAGGTTCTGCGCAG 663
 QY 829 AGGATTTGAAGAGAGATGTCGCCAACAGTCTTCTGATGAGATTAATTTCTGTTG 888
 DB 664 CGTATTATCAACGCGAAGTGCAGGAGGTTGAAAGCGCGCGGCTACTGGCGCTG 723
 QY 889 ATGGGTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
 DB 724 ATGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 783

466 GAAGTAAACTATGAAGCTTTAGAAAAATATGGTCTGTGATCTTTGTAGAGAAGTACGTCAA 525
688 GGAAGAGCTTGATCCCTGTGATTTGGTCTGTGATGAGGAGATAGAGAGTCTGTGAGATCTTT 747
526 GGTAAATGGAACCTCTGTATCGGTAGATGAGGAAATACGTAATACAAATTCGTATATTA 585
748 TCGAGGAGAACCAAGCAATCCTGTGCTTATTTGAGAGAGCCAGGAGTTGGTAAACAGCT 807
586 AGTCGAAAACAAAATAATCCGGTTTATCCGTGAACCCAGGTGTAGGTAAACAGCA 645
808 GTGGTTGAAGGTTTACGCAAAAGGATTTGAAAGAGATGTGCCCAACAGTCTTACTGAT 867
646 ATTGTTGAAGGTTTACGCAACCAATTTGACGTAAGAGCGTTTCCAGAAATCAATCTCGAT 705
868 GTGAGATTAATTTGTTGACATCGGTGGTGTAGTTGCTGGTGTGCTAAATACCGAGGAG 927
706 AAAACGATTTTCAGCTTGATTTAAGTGTCTTAGTTGCAAGGTGCAAAATATAGAGTGA 765
928 TTTGAAGAAAGGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGAAGGCCAAAGTGAT 987
766 TTTGAAGAAAGGTTTAAAGCAGTATTTGAAGAAAGTAAAGAAATCTGAAGGTGCAATATC 825
988 CTCTTTATTTGATGAGATTCATTTGGTCTTGTGCTGTGCAAAACGTAAGGGTCTGATGGAT 1047
826 TTATTCATTTGATGAAATTCATATGTAGTAGCGCTGTGTAAGCTGATGGTCTATGGAC 885
1048 GCAGCTTAATCTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCAATGCTGCTACA 1107
886 GCAGGAAACATGTTGAAACCAATGCTTGGCGGTGGTGAATTCGATTTGATAGTCTACG 945
1108 ACCGTTTGAAGATACAGGAAATATGTTGAGAAGATGCTGCTTTGAGAGGAGGTTCCAA 1167
945 ACATTAATTAATCTGATGATATATTTGAAGAAAGATTCAGCAATTAAGAGCTGCTTCAA 1005
1168 CAAAGTCTATGTTCCGAGACCAAGTGTGCTGTGACCAATTTAGTATCCTTAGAGAGCTCAG 1227
1006 AAGTTTGGGTCAGTGAACCTGATGTTGAGAATACAAATATCAATTTTACGTGCTTAA 1065
1228 GAGAATGATGAGGACATCATGTTGCGCAATCCAGACAGAGCTCTTATAAATGCTGCT 1287
1066 GAACGTTATGAGGCTTACATGTTGACGATGATTTTCCAGATAAAGCTATAGATTTAGTC 1125
1288 CAGCTGCTGCTGTTACATAAATGTTGCGCAATTTACCGATTAAGCAATGATTTGGTT 1347
1126 GAGCTATCTGATAGATATATTTCCGATAGATTTTCCAGATAAAGCTATAGATTTAGTC 1185
1348 GATGAGGCTTGTCCGAATGTGAGAGTCCAGCTGTAGTCAACCTGAAGAGATGATATAC 1407
1186 GACCAAGCTTGTGCGCAATACGTA CAGAAATGGGATCAAAATCCAAAGAGTTAGACCAA 1245
1408 CTTGAAAGGAGAGGATGACGCTGGAATTTGAATTTCAAGCTTGAAGAGGAGAGAT 1467
1246 GTAAATCTGCTGATGATGATGATTTAGAAATTTGAAGAAAGCGCTTTGAATGAATCTGAT 1305
1468 AAGCCAGCAAGCTGCACTTATAGAGTGTGCGAAAGAGCTTGAATGACCTGAGAGCAAG 1527
1306 AAGCAAGTAAACATCGTTTGAAGAGCTTCAAGAGGAAATTTATCTAATGAAAGAAAA 1365
1528 CTTGAGGCTCTCAGATGAATACAGAAAGGAGAGAGATTTGATGATTTGCAAGG 1587
1366 CAATCTCTTTAAATCAAGTGTGAAACAGAAAGAAAAATTTGTAAGTACAGAA 1425
1588 CTTAAACAGAAAGAGAGAGCTCATGTTTCTTTGAGAGGAGGAGCAACGAGATATGAC 1647
1426 AACGCTGCTGAACTTTGATGATGATGCTGCTGAGCTTTGAGATGCAACAGAGGTAAT 1485
1648 CTTGCAAGAGCTGCTGATCTAGATATGAGGCAATTCAGAGTGGAAATCTGCAATGGC 1707
1486 CTTGAAAAAGCAGCAGAAATTAACATATGTTAGCAATTCGCAATTTAGAAAAAGAAATTCGCA 1545
1708 CAACCTGAGAGAACTTCTCTGAGAGAAATG-----TGATGCTCAGAGAA 1752
1546 GAATTTGAGGAAGCTTTCAAGATGAAGAACTGGCGAAGATAGTGAAGAAATGATCTGTGAA 1605

QY 1753 AACGTTGGGCTGAAACACATTTGCTGAGGTTTGAGCCGTTGGACAGGATTCAGTGACG 1812
Db 1606 GTGTTTCAGATGAAGAAATAGCGACATAGTTAGTCAATGGACAGTATACCTGTATCT 1665
QY 1813 AGACTTGGCCAAATGAGAGAGAGAGGTTGATTTGGTCTTGTGATAGGTTGCATAGCGG 1872
Db 1666 AAACTTTGTGAAACAGAAAGAGAAAAAGCTATTAAAGTTTAAAGTGACATCTTACAAACGT 1725
QY 1873 GTTGTGGGACAGAAATCAAGCGTAAATGCAAGTTTCTGAGGCAATCTAAGGTCAAGGGCA 1932
Db 1726 GTAGTAGGTCAGATTAAGCAGTTGATTTAGTATCAGACGAGTAGTTAGAGCACGTGCT 1785
QY 1933 GGAATTTGGTGGGACACAAACAGCCAACTGGATCATCTTATCTCTTGGACCAACTGGTGT 1992
Db 1786 GGAATTAAGATCCGAATAGACCAATCGGAAGTTTCTTATCTTTAGGACCTACTGGAGTA 1845
QY 1993 GGCATAACTGAGCTCCCGAGGCTCTTGTGAGCAGCTGTTGATGATGAAAACTCTTA 2052
Db 1846 GGTAAACTGAATTTAGCAAAATCGCTTGTCTTCACTTTTTCGATTTCTGAAAAACATATG 1905
QY 2053 GTTCGATTTGATATGTCGAAATATATGGAACAAACATCTCTCTCTCGCCTCATTTGGGCA 2112
Db 1906 ATTAGAATTTGATGAGCGAATATATGGAACAAACATCTCTCTCTCATCTTATTTGGTGA 1965
QY 2113 CCACAGGTTATGTTGGTCAAGAGGAGGTTGGAACAATCTAGCTGAGGCTGTGAGGAGCGA 2172
Db 1966 CCTCCAGTTATGTAGGTCACGATGAGGTGGTCAATTAATCTGAAGCAGTTAGACGTAAT 2025
QY 2173 CCTTATTTGCTCATCTCTTGTGATGAGTGAGAGGCTCATCTGCTGCTTCAACACT 2232
Db 2026 CCATCTCAGTTATTTTGTGATGAGGAGGTTGAAAGACATAGCAGATGTTTTTATGTA 2085
QY 2233 CTGCTCCAAGTTTGGATGATGTTGATGAGGAGGCTGATGAGGAGGAGGAGGAGGAGG 2292
Db 2086 TTACTTCAATCTAGATGAGGTGGTCTTACGGAATCTTAAAGTAGAGGTTTGAATTT 2145
QY 2293 AGGAATCTGGTGAATCATGACATCAACCTTGTGCTGAAACCTCTTGCAGGCTA 2352
Db 2146 AAAAATACCATTTATCATCATGATAGTAATTTTGGTTTCACAAGTATTTACTTGAATGTA 2205
QY 2353 A---CTGGGAAGTAAACAAATGGAAGTGGCGGAGCTGTTGATGCGGAGGAGGAGGAAA 2409
Db 2206 AAGATGCTGTTGAAATTTAGTATGATACAGAGAAAGCAGTTATGACAGCTCTACATGCA 2265
QY 2410 CACTTCAGACAGAGCTCTTTGAAACAGGCTTGGAGAGATTTGGTGTTCGACCCCTTCA 2469
Db 2266 TACTTCAACCTGAAATTTAAATCGTATGATGATGATCGTGTATTTTAAACCATTTATCA 2325
QY 2470 CATGACCAAGTTGAGGAAAGTAGCTCGGCTTCAATGAAGACGTTGCTGCTCGGCTGCT 2529
Db 2326 GTTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2385
QY 2530 GAAAGAGAGGTTGCTTTGGGAGTCACTGATGCTCTTGTGAGCTATATCTTGGCAGAGAT 2589
Db 2386 GATCAACATATCTCAATTTGAAGTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2445
QY 2590 TATGACCCGGTGTATGTTGCTGAGCCCTATAGAGAGATGATGAGGAGAGAGAGGTTGAC 2649
Db 2446 TATGAACCAAAATTTGGTGCAAGACCAATTAAGAGCTTTGTTTCAACGACAAATAGAA 2505
QY 2650 GACTGTCAAGAGTGTGTTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2709
Db 2506 CCAATTCAGCTATGATGATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2565

RESULT 8

US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention of infection
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention of infection
 TITLE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 CURRENT FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 Q ID NO 1
 LENGTH: 1230025
 TYPE: DNA
 ORGANISM: Chlamydia pneumoniae
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)..(15000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (15001)..(30000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (30001)..(45000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (45001)..(60000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (60001)..(75000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (75001)..(90000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (90001)..(105000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (105001)..(120000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (120001)..(135000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (135001)..(150000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (150001)..(165000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (165001)..(180000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (180001)..(195000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (195001)..(210000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (210001)..(225000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (225001)..(240000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (240001)..(255000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (255001)..(270000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (270001)..(285000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (285001)..(300000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (300001)..(315000)
 OTHER INFORMATION: n=a or c or g or t

2005	CTCGCCAAAGGCTCTTGCTGAGCAGCTGTTTGATGATGAAGAAACCTCTTAGTTCGGATTGAT	2064
190175	CTTGC AAAAGCTCTTCAGATCTTCTTTTCAATTAAGAGGAAGCTATGTCGCTTCGAT	190234
2065	ATGTCCGAATATATGGAACAACACTCTGTCTCTCGCTCATTTGGGGCACACACAGGGTAT	2124
190235	ATGTCAGAGTATATGGAAGAAGCATTCATTTCCAAGCTTATAGGATCTTCTCCAGGGTAT	190294
2125	GTTGGTCCAGAGGAGGTGAGCAACATTAACGTAGGCTGTGAGGAGGCGACCTTATTGTGTC	2184
190295	GTGGGTTTATGAGAAGGTGGAGTCTTTCTGAGGCTCTTCGACGAGCTCCCTATTTCAGTA	190354
2185	ATACTCTTTGATGAAGTGAAGAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAAGTT	2244
190355	GTTCTCTTTGATGAGATAGAGAAGCAGATAGGAAGTTCCTAAATATCCTTTTACAGTT	190414
2245	TTGGATGATGGTCGATTTGACAGACGGCAGGACAGTCTGATTTACGGAACTCGGTG	2304
190415	TTTGATGATGGGATTTCTTACGGATGGAAAAACGAAAGTAAATTTGTAATAATGCTTGT	190474
2305	ATAATCATGACATCAAACTTGCTGTCTGACACCT-----CCTTGCAGGGCTTACT	2355
190475	TTTATCAAGACATCAATATAGGTTCTCCAGAACTTGTGAGATTTATGTTTCAAAAAAGGA	190534
2356	GGGAAAGTAAACAATGGAAGTGGCCGGGACTGTGTGATCGGAGGTGAGGAAACACTTC	2415
190535	AGTGAGCTTACGAAAGAAGCGATTCTTTCTGTAGTCTCTCCAGTATTGAAGAGTACTTG	190594
2416	AGACCCAGAGCTCTTGAACAGGCTTGAGAGATTGTGGTGTTCGACCCCTTTCACATGAC	2475
190595	AGCCCTGAATTTATGAACCGAATTGATGAGATATCTCTCTTTGTTCCATTAAACGAAAGAA	190654
2476	CAGTTGAGGAAAGTAGCTCGGCTTCAAATGAAAGAGTTGCTGTCCGGCTTGTCTCAAGA	2535
190655	GATATCGTGAATAGTTGSCATTCAAATCGGAAGATTGCCAGAGATTAAAGCACGG	190714
2536	GGAGTTGCTTTGGCAGTCACTGATGCTGTTTGGACTATATCTTGGCAGAGATTATGAC	2595
190715	CGGATCAATTTATCTTTGGGATGATTCTGTAATATTTATTTCTTAGTGAAACAGAGTTATGAC	190774
2596	CCGGTGTATGGTCTAGGCTATAAGGAGATGGATGGAGAAGAGTGTGTGACAGACTG	2655
190775	AGTGTCTTCCGAGCCGCCCTTTAAACGTTTGTATCCCAAAAAAGTTGTGATCTTGCTT	190834
2656	TCAAG 2661	
190835	TCCTAAG 190840	

LT 9
9-199-637A-280
quence 280, Application US/09199637A
tent No. 6355411

CELL NO. 6355411
NERAL INFORMATION.

GENERAL INFORMATION: Ausubel
 APPLICANT: Ausubel Frederick

APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M

APPLICANT: GOODMAN, HOWARD M.
APPLICANT: RAHME, LAWRENCE G.

APPLICANT: KAHME, DAUNCE G.
APPLICANT: Mahajan-Miklos, Shalina

APPLICANT: Manojan-Maxios, Sinc
APPLICANT: Tan. Man-Wah

APPLICANT: Cao, Hui

APPLICANT: Drenkard, Eliana

APPLICANT: Tsongalis, John

TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

TITLE OF INVENTION: SEQUENCES AND USES THEREOF

FILE REFERENCE: 00786/361002

CURRENT APPLICATION NUMBER: US/09/199,637A

CURRENT FILING DATE: 1998-11-25

RIOR APPLICATION NUMBER: 60/066,517

PRIOR FILING DATE: 1997-11-25

NUMBER OF SEQ ID NOS: 437

SOFTWARE: FALCON

Q	ID	NO	280
T	ENCTH-	2590	

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-280

```
Query Match      17.6%; Score 546.8; DB 4; Length 2580;
Best Local Similarity 53.8%; Pred. No. 4.5e-148;
Matches 1158; Conservative 0; Mismatches 987; Indels 9; Gaps 1;
```

530	TCAGGGATTTTGTTGAACGAAGTTCGGTGTAGCGCGCGAGGGTTAAAGTCTGAGGTTGAGA	589
380	TCCGCAAGCTGCTGCTCGCCACAGGGCGTTCGCGCAAGCGCTTGGAGAATGCCGTGGCCA	439
590	AGCTTCGTGGGAAAGAAAGATTGAGAGTGCCTTCAGGGGACACAAATTTTCAAG	649
440	ACTTGGTGGCGGAGCGGTGAACGCCGAAAGTCGAGGAGTCCGCCAGGCGCTGG	499
650	CTTTAAAGACTTATGGAAGAGATTTGGTTTGACCAAGCAGGGAAGCTTGATCTCTGTGATTG	709
500	ACAAGTACACCGTCGACATGACCAAGCGCGCGAGGAAGCGAAGCTCGACCCGGTATCG	559
710	GTCTGTATGAGAGATTAGAGAGTGTGAGGATTTCCTTCGAGGAGAGACGAGAACATTC	769
560	GTCCGGACACAGAGATCCGCCGAGCCATCAGGTCTCGAGCGGGGACCAAGAAACAAC	619
770	CTGTGCTTATTCGAGAGCCAGAGTTGGTGTAAACAGCTGTGCTTGAAGTTTAGCACAA	829
620	CGGTGCTGATCGCGCAACCGCGCTCGGACAGCCGCATCGTCAGGGGCTTGCGCCAGC	679
830	GGATTGTGAAGGAGATGTGCCCAACAGCTCTTACTGATGTGAGATTAAATTCGTTGGACA	889
680	GCATCATCAACGGCGAAGTCCGGACGGCTCAAGSACAAGCGCTGCTGGCCCTGGACA	739
890	TGGGTGCGTTAGTTCGTGCTGCTAAATACCGAGGAGAGTTTGAAGAAAGTTGAAATCTG	949
740	TGGGGGCGCTGATCCGCCGTGCCAAGTTCCCGCGCGAGTTCGAGGAAACGCTTGAAGCGG	799
950	TTTTCAAGAAGTTCGAGGAGCTGAAGGCAAGTCAATCTCTTTATTCATGAGATTCAAT	1009
800	TCCTCAAAGACTGGGCAAGCAAGGAAGCGCGGTCAATCTGTTCACTGACGAACTGCA	859
1010	TGGTTCTTTGTCCTGGCAAACTGAAGGATCGATGGATGCAGCTAATCTGTTCAAGCCCA	1069
860	CCATGGTCGGCGCGCAAGCGGAAGGTGCCATGGACGCGCGCAACATGCTCAAGCGG	919
1070	TGTTAGCTAGAGGGCAGCTTCGATGCATTTGGTGTACAAAGCTTGAAGAAATACAGGAAT	1129
920	CTCTCGCGCGGGCAGCTGCACTCGCTGGTGTACTACCTCGACGAGTATCGCCAGT	979
1130	ATGTTGAGAAAGATCTGCTTTGAGAGGAGGTTCCAAAGTCTATGTTGGGAGCCAA	1189
980	ACATCGAAGAGATGCGCGGCTGGAGCGCGCTTCAGAGGTGCTGGTGGACGAAACCGA	1039
1190	GTGTGCTTGACACCAATTAGTATCTTTAGAGACTCAAGAGGAAGTATGAGGGACATCATG	1249
1040	GCGAGGAGAACACCATCGCCATCTCTCCGTGGCTCAAGSAAACGCTATGAAGTGCCACC	1099
1250	GTGTGCGAATCCAAGACAGCTCTTAATAATGTCTGCTCAGCTGTCTCTCGTTACATAA	1309
1100	GGGTGAGCATCACGACCGCGCGATCATCGCGCGGCCAAGCTGTGCACCGCTACATCA	1159
1310	CTGTGCGGCATTTACCGGATAAAGCAATTGATTTGGTTGATGAGGCTTGTGCAATGTGA	1369
1160	CCGATCGGCACTGCGGACAAAGGCCATCGACCTGATCGACGAGGCGCCAGCCGATCC	1219
1370	GAGTCCAGCTTGATGTCAACTGAGAGATTGATTAACCTTTGAAAGGAAGGATGCGAGC	1429
1220	GCAATGGAGATCGACTCCAAAGCGGAGGAATCGGATCGTCTCGACCGCTGATCCAGC	1279
1430	TGGAAATTAAGTCTCAGCCTTGGAAAGGAGAGGATAAGCCAGCAAGAGCTCGACTTA	1489
1280	TGAAGATCGACCGGAGCGCTGTGAAGAAGGAAGACGACGAGGCCACCAAGGAAGCGCTGG	1339
1490	TAGAGGTTCGGAAGAGCTTGATGACCTTGAGAGCAAGCTTTCAGCCTCTCACGATGAAT	1549

1340 CCAAGCTGGAGGAGATATCGTCAAGCTCGAGCGCAATACGCGACCTCGAGGAGATCT 1399
1550 ACAGAAAGGAGAGAGAGATTTGATGAGATTCGAAGCTTTAAACAGAAAGAGAGAGC 1609
1400 GGAAGTCGAG 1459
1610 TCATGTTTCTTTTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669
1460 CCAAGCAGGAGATGAG 1519
1670 GATATGGCGCAATCAAGAGTGGAAATCTGCAATTCGCCAACTTGAAGGAATCTTTCTG 1729
1520 AGTAGCAGAGACCATCCCGACCTTGAACGAGCGCTGCAGATGCTCACCAGCAGCGCA 1579
1730 AAGAGATGTGATGCTCAGAGAAAGCTTGGCCCTGGAACATCTGAGTGTGTGAGCC 1789
1580 CCGAGAACGAGTTGCTGCGCAACAAGGTGACGACGAGGAATCCCGAAGTGTTCCTCA 1639
1790 GTTGACAGGAGATTCAGTAGCAGAGATTTGGCCAAATGAGAAGAGAGAGTGTGATGTC 1849
1640 AGTGACCGGTATCCCGGTGTGGAAGATGCTGAGCGGAGCGGAGAGAGTGTGCGCA 1699
1850 TTGCTGATAGTTGATAG 1909
1700 TGGAGCAGAGAGTGCATCGCGAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1759
1910 AGGCAATTTCAAGTTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969
1760 AGCGGTGGCGGTTCGCGCGCGCGCTCGCGATCCGAAACCGCGCGAGCGCTGTTC 1819
1970 TATCTTTGGACCAACTGGTGTGGCAAACTGAGCTCGCAAGGCTCTTCTGAGCAGC 2029
1820 TCTCTCTCGCGCGAG 1879
2030 TGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2089
1880 TCTTCGATACGAG 1939
2090 CTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2149
1940 CGGTGGCGGCTGATCGCGCGCTCTCGCGCTACGTCGGCTTCAGAGAGAGAGAGAG 1999
2150 TAACTGAGCTGTGAG 2209
2000 TGACGAG 2059
2210 CTGATGTTGCTCTCTCAACTCTGCTCAAGTTCGCTCAAGTTCGCTCAAGTTCGCT 2269
2060 CCAATCCGAGATGATTTCAACATTTCTCTCCAGGTGCTCGAGAGAGAGAGAGAGAG 2119
2270 GGCAAGGAG 2329
2120 GTCAAGGAG 2179
2330 CTGAACACCTCTCTGAG 2389
2180 CGGCGCAGATCCAGG-----AGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2230
2390 TGATCGGAG 2449
2231 TGATGAG 2280
2450 TGTGTTTCAG 2509
2291 TGTGTTTCAG 2340
2510 AGTGTGCTCTGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2569
2351 GCTTGGCAAGGCGCTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2410
2570 ACTATATCTTGGCAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2629

2411 ACAAGCTGATTCGCGTTCGAGCCCGCTTATGCGCAAGCCCGCTGTAAGCGGCGCA 2470
2630 TGCAGAGAGAGTGTGACAGAACTGCTCAAGAGTGTGCTGCGTGAAGAAATCG 2683
2471 TCAGCGCTGATTCGAGAAACCCGCTGCGCAACTGATCTGCGCGCAAAATTCG 2524

RESULT 10
US-09-199-637A-272/c
; Sequence 272, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199, 637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066, 517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-272

Query Match 17.6%; Score 546.8; DB 4; Length 2970;
Best Local Similarity 53.8%; Pred. No. 4.8e-148;
Matches 1158; Conservative 0; Mismatches 987; Indels 9; Gaps 1;

530 TCAGGAGATTTGTTGAACGAGTGGGTGAGCGAGCGGAGGTTAAAGTCTGAGGTTGACA 589
2585 TCAGGAGATTTGTTGAGCGAGGCGGTGTCGCGCGAGGCGGTGCGCGAGGAGATCGCGTGGCCA 2526
590 AGCTTCGTGGGAAAGAGGAGAAAGTTGAGAGTCTTCAGGGGACACAAAATTTTCAAG 649
2525 ACCTGCTGGCGGAGAGCGGTGAACGACCGGAGAGTTCGAGGAGTTCGCGCGAGGCTGG 2466
650 CTTTAAGACTTATGAGAGAGATTTGTTGAGCAAGCAGCAGGAGCTTGATCTGTTGATTTG 709
2465 ACAAGTACACCGTGCAGATGACCAAGCGCGCGAGGAGGCAAGCTCGACCGGTTGATCG 2406
710 GTGCTGATGAGGAGATTAGAGAGTCTGAGGAGTCTTTTCAGAGGAGAAACGAGAAACAATC 769
2405 GTGCGAGCAGAGATCCCGGACCATCCAGTCTGCGAGCGGCGGACCAAGAGAACCC 2346
770 CTGTGCTTATTCGAGAGCGAGTTGGTAAACAGCTGTGTTGAGGTTTTCAGCAAAA 829
2345 CGGTGCTGATTCGCGGAAACCGCGGTTCGCGAGAGACCCATCTGTCGAGGCGCTGGCGCCAG 2286
830 GGATTTGAGAGAGATGTCGCCAACAGCTTACTGATGATGAGATTAAATTTCTGTTGACA 889
2285 GCATCATCAACCGCGAGTGCAGGAGTCCCGGCTCAAGGCAAGCGCTGCTGGCGCTGGACA 2226
890 TGGGTGCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 949
2225 TGGGGCGCTGATCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2166
950 TTTTGAAGAGAGTTGAGGAGCGCTGAGAGGCAAGTGTCTCTTTATGATGAGATTCATT 1009
2165 TCCTCAACGAACCTGGGCAAGCAGGAGGCGGCTCATCTGTTTCATCGAGCAACTGCACA 2106
1010 TGGTTCTTGTGCTGCGCAAAACCTGAAGGCTGATGAGTGCAGCTAATCTGTTCAAGCCCA 1069

2150	TA	ACTGAGCGCTGAGGAGGCGACCTTATTGTGTCATACCTTTTGTATGAAGTGGAGAAGG	2209
965	Db	TGACCGAGGCGATCCGCGCAGCCCTACTCCGGTGGTGTCTGCTGCGACGAGGTGGAGAAGG	906
2210	QY	CTCATGTGTGTCGCTTTCAACACATCTGCTCCAAAGTTTTGGATGATGTTGCAATTCACAGACG	2269
905	Db	CCCATCCGGATGTATTCAACATCTCTCTCCAGGTGCTCGAGGACGAGCGCTGACCGACA	846
2270	QY	GGCAAGCGAGGACAGTCGATTTTCAGGAACTCGGTTGATAATCATGACATCAAACTTTGGTGTG	2329
845	Db	GTACAGGGCGTACGGTGGACTCTCCGCAACACCGTGTGTGTGATGACTCTCCAACTCGGTT	786
2330	QY	CTGAACACACTCTCTTGACGGCTAACTGGGAAAGTAACAATGGAAGTGGCCCGGACTGTG	2389
785	Db	CGSCGCAGATCCAGG-----AGCTGCCGCGACCCGCGAGGCGCAACGTGCCGCG	735
2390	QY	TGATCGGAGGTGAGGAAACACTTCAGACGAGAGCTTTTGAAACAGGCTTGACGAGATTG	2449
734	Db	TGATGACCGGCTCAATGCGCACTTCGTCGCGAATTCATCAACCGGATTCGACGAAGTGG	675
2450	QY	TGCTGTTTCGACCCCTTTTCACATGACCACTGTGAGGAAAGTAGCTCGCTGCTCAAAATGAAG	2509
674	Db	TGGTGTTCGAGCGCTGGCTTCGGAGCAATCGCCGGCATCGCCGAGATCCAGCTCGTTC	615
2510	QY	ACGTGCTCTCGCGCTTGCTGAAAGAGGAGTTGCTTTTGGCACTACTGATGCTGCTTTGG	2569
614	Db	GCTTCGCGAAGGCGCTTGGCCGAGCGCGAGCTGAGCCTCGAACTTGAGCCAGGAGCGCTCG	555
2570	QY	ACTATATCTTTGGCAGAGAGTTATGACCCCGGTGTATGGTCTTAGGCCTATTAAGGAGATGA	2629
554	Db	ACAAGCTGATTCCGCTCGGCTTCGACCCCGGTCTATGGCGCACGCCCGCTGAAGCGGGCCA	495
2630	QY	TTGGAGAAGAGTGGTGTGACAGAACTGTCTCAAGATGTTGTGGTGAAGGAAATCG	2683
494	Db	TCACGCGTGTGATCGAAGACCCGCTTGGCGCACTGATCCTGCGCGCAATTCG	441

US-09-199-637A-1/c

```
Query Match      17.6%; Score 546.8; DB 4; Length 42235;
Best Local Similarity 53.8%; Pred. No. 2.2e-147;
Matches 1158; Conservative 0; Mismatches 987; Indels 9 Gaps 1;
```

530 TCAGGATTTGTTGAACGAAGTCGGTGTAGCAGCGCGAGGGTAAAGTCTGAGGTTGAGA 589
25812 TCGGCAAGCTGCTGCTCGCCAGGCGGTCTCGCGAAGCGCTCGAGAAATGCGTGGCCA 25753
590 AGCTTCGTGGGAAAGAGGAGAAAGTTGAGAGTGTTCAGGGAGACAAATTTTCAAG 649
25752 AACTGCTGTGGCGGAGCGGTGAACGACCCGAACGTCGAGAGTTCGCGCCAGCGCGTGG 25693
650 CTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTTGATCTGTGATTG 709
25692 ACNAGTACACCGTGCATGATCAAGAGCGCGCGAGGAGGCAAGCTCGACCCGTTGATCG 25633
710 GTGCTGATCAGGAGATTAGAAGAGTGTGAGGATTTCTTTCGAGGAGAAAGCAAAATC 769
25632 GTCCGACGACGAGATCCCGCCGACCATCCAGGTCTCGCAGCGCGGACCAAGAACCAAC 25573
770 CTGTGCTTATTGAGAGCCAGAGTTGTTAAACAGCTGTGTGTAAGTTTGAACAAA 829
25572 CGTGTGCTGATCGGCGAACCCTCGTGGCAAGCCGATCGTCGAGGGCTTGGCCACG 25513
830 GGATTCGAAAGAGATGTGCCAACAGTCTTACTGATGTGAGATTTAAATTCGTTGGACA 889
25512 GCATCATCAACCGGAGTTCGCGAGCGCTCAAGGACAAAGCGCTGTGCGCTGGACA 25453
890 TGGGTGCTGTAGTGTGCTGTTAAATACCGAGGAGGTTTGAAGAAAGTTTGAATCTG 949
25452 TGGGGCGCTGATCGCGGTGCCAAGTTCCGCGCGGAGTTCCGAGGAACGCTGAGGCGG 25393
950 TTTTGAAGAAGTTGAGGACGCTGAGGCAAGATGATCTCTTTTATGATGATGATTCATT 1009
25392 TCTTCAACGAACTGGGCAAGCGAGGAGGCGGGTCACTGTTTATCAGCAAACTGACAC 25333
1010 TGGTCTTGGTCTGCGCAAACTGAAGGTCGATGATGAGTCACTAATCTGTTCAAGCCCA 1069
25332 CCAATGTCGCGCGCGCAAGCGGAGGTCATGAGCGCGCAACATGCTCAAGCCGG 25273
1070 TGTAGCTAGAGGCGAGCTTCGATGCAATGTTGTTGCTTAAACGCTTGAAGAAATACAGGAAT 1129
25272 CTCTGCGCGCGCGAGCTGCACTCGTGTGTTGCTACTACCTCGACGAGTATCGCCAGT 25213
1130 ATGTTGAGAAAGATGCTGCTTTGAGAGGAGTTCACAAAGTCTATGTTGCGGACCA 1189
25212 ACATCGAGAGAGATCGCGCTGAGAGCGCGCTTCAGAGGTGCTGTGAGCAACCGA 25153
1190 GTGCTGCTGACACCAATAGTATCTTTAGAGGACTCAAGGAGGATATGAGGACATCATG 1249
25152 GCGAGGAGACACCATCGGCATCTCTCGTGGCTCAAGAAAGCTATGAAGTGACACACG 25093
1250 GTGCTGATCCAGACAGCTCTTATAATGCTGCTCAGCTGTCTGCTCGTTACATAA 1309
25092 GGGTGAATCATCCAGCGCGGATCATCGCGCGGCAAGTGTGCAACCGCTACATCA 25033
1310 CTGGTCCGATTTACCGGATAAGCAATGATTTGTTGATGAGGCTTTGCGAATGTA 1369
25032 CCGATCGCACTCGCGACAGGCAATCGACTGATCGACGAGCGCGCCAGCCGATCC 24973
1370 GAGTCCAGCTTGATAGTAACTGTAAGAGATTGATACTTGAAGAGGAGGATGACG 1429
24972 GCATGAGATCGACTCAAGCGGAGGAACTGGATCGTCTCGACCGTCTGATCCAGC 24913
1430 TGGAAATTTGAATTCAGCCCTTGGAAAGGAGAGGATTAAGCCAGCAAGCTCGACTTA 1489
24912 TGAAGATCGAGCGGAGGCTGGAAGAGGAGAGCGAGAGCCACCAAGGAGCGCCCTGG 24953
1490 TAGAGTCCGGAAGAGGCTTGATGACCTGAGAGACAAGCTTCAGGCTCTCAGATGAAAT 1549
24852 CCNAGCTGAGGAGGATATCGTCAAGCTCGACGCGGATACGCCGACCTCGAGGAGTCT 24793
1550 ACAGAAAGAGAGAGAAATGATGATTCGAAGGCTTAAACAGAAAGAGAGAGC 1609
24792 GGAAGTCCGAGAGGCGGAGGTGCGAGGCTCGCGCAGATCCAGCAGAGATCCGAGCAGG 24733

QY 1610 TCATGTTTTCTTTTCAGAGGCGAGAACGAGATATGACCTTTGCAAGAGCTGTGATCTAA 1669
Db 24732 CCAAGCAGGAGATGGAGCGCGCGCGCAAGCGGCACTTCGAGAGCATGGCGCATCC 24673
QY 1670 GATATGCGCAATTTCAAGAGTGGAAATCTCAATTTGCCCACTTGAAGNACTTCTCTG 1729
Db 24672 AGTACCAAGACATCCCGGACCTGGAAACGACGACCTGAGATGGTTCGACCGCAAGA 24613
QY 1730 AAGAGAAATGTGATGCTCACAGAAAACTTTGGGCTGAAACATTTGCTGAGGTTGTGAGCC 1789
Db 24612 CCGAGAACAGTTGCTGCGCAACAAAGTGTGACCGAGGAAATCGCGAGTGTGTTCCA 24553
QY 1790 GTTGCACAGGATTCAGTGTGACGAGACTTTGGCCAAATAGAGAGGAGGTTGATGGTC 1849
Db 24552 AGTGCACCGGTATCCCGTGTGGAAGATGTCGAGGCGAGCGGAGAGCTGCTCGCA 24493
QY 1850 TTGCTGATAGTGTGATAGCGGGTTGTGGGACAGAAATCAAGCGGTAATGCAAGTTCTG 1909
Db 24492 TGGAGCAGGAGCTGCAATCGCGGAGTGTATCGGCCAGGACGAGGCGGTAGTCGCGGTCCA 24433
QY 1910 AGGCAATTTCTAAGTCAAGGGCAGGACTTTGGTAGGCAACAAAGCAAACTGGATCAATCT 1969
Db 24432 AGCGGTGCGCGTTCGCGCGCGCGCTCGCCGATCGAAACCGCGAGCGGTCTGTTCC 24373
QY 1970 TATTCCTTGGACCAACTGTGTGTTGGCAAACTGAGCTCGCCAAAGGCTCTTCTGAGCAGC 2029
Db 24372 TCTTCTCGGCGGACCGCGGGTGGGCAAGACCGAGTTGTGCAAGGCGCTGCGCGAGTTCC 24313
QY 2030 TGTTCGATGATGAAACCTTTAGTTTCGGATTTGATATGTCGGAATATATGGAACACACT 2089
Db 24312 TCTTCGATACCGAGGAGCGCTGTCGCGATATATGTCGAGTTTCATGAGAAACACT 24253
QY 2090 CTGTCTCTCCCTCATTTGGGGCACCAAGGTTATGTTGTCACAGGAGGTTGGACAAC 2149
Db 24252 CGGTGCGCGCTGATCGCGCGCTCGGCTACGTGCGCTTCGAGGAGGCGGCTACC 24193
QY 2150 TAACTGAGGCTGTGAGAGGCGACTTTTGTGTATCTTCTTTGATGAGTGGACAAG 2209
Db 24192 TGACCGAGGCGATCCGCGCAAGCCCTACTCTCGTGTGCTGTCGAGAGGTTGGAGAAG 24133
QY 2210 CTATGTTGCTGCTTCAACACTCTGCTCCAGTTTGGATGATGTCGATGACAGAG 2269
Db 24132 CCATCGGATGATTTCAACATTTCTCTCCAGTGTCTGAGGACGACGCTGACCGACA 24073
QY 2270 GGCAGGACGAGCAGTCGATTTTCAGAACTCGGTGATAATCATGACATCAAACTTTGGTG 2329
Db 24072 GTCAAGGCGGTACGGTGGACTTCGCAACACCGTGTGTTGATGATCTCCAACTCGGTT 24013
QY 2330 CTGAACACCTCTTTCAGGCGTAACTGGGAAAGTAACATGGAAGTGGCCCGGACTGTG 2389
Db 24012 CGCGCAGATCCAGG-----AGCTGGCGGCGAACCGGAGGCGAACGTCGCGCAG 23962
QY 2390 TGATCGGAGGAGTGGAGAAACACTTCAGACCCAGAGCTTTTGAACAGGCTTGACGATTTG 2449
Db 23961 TGATGAGCGGCTCAATGCGCACTTCGCTCGGAATTCATCAACCGATCGACGAGTGG 23902
QY 2450 TGGTGTTCGACCCCTTTTCAATGACATGACGATGAGGAAAGTAGCTGCTTCAATGAAG 2509
Db 23901 TGGTGTTCGAGCGGCTGGCTCGGAGCAGATCGCCGCAATCGCCGAGATCCAGCTCGGTC 23842
QY 2510 ACCTTGTGTCGCGCTTGTGAAAGAGGAGTTGCTTTTGGCAGTCACTGATGCTGTTTGG 2569
Db 23841 GCCTGCGCAAGCGGCTGGCGGAGCTGAGCCTGGAATCTGAGCCAGGAGGCGCTGG 23782
QY 2570 ACTATATCTTGGCAGAGATTATGACCCGCTGTATGGTGTAGGCTTAAAGAGATGGA 2629
Db 23781 ACAAGTGTATTCGCTCGGCTTCGACCCCGTCTATGCGCACGCCGCTGAGCGGCCA 23722
QY 2630 TGGAGAGAGAGGTTGTGACAGAACTCTCAAGATGTTGTGCTGAGGAAATCG 2683
Db 23721 TCCAGGCTGGATCGAGAACCCGCTGCGCAACTGATCTCTGCGCGCAAAATTCG 23668

JLT 12

09-252-991A-4842
Sequence 4842, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 4842

LENGTH: 2580

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

09-252-991A-4842

Query March 17, 64; Score 545.2; DB 4; Length 2580;
Best Local Similarity 53.7%; Pred. No. 1.3e-147;
Matches 1157; Conservative 0; Mismatches 988; Indels 9; Gaps 1;

530 TCAGGGATTGTTCAACGAAGTCGGTGTAGCGCGCGAGGTTAAAGTCTGAGGTTGAGA 589
380 TCAGCAAGCTGCTCTCGCCAGCGCGGTGTGCGCAAGGCGCTGGAGAAATGCGGTGGCCA 439
590 AGCTTCGTGGGAAAGGAAGGAAGGTTGAGAGTGTCTCAGGGGACACAAATTTTCAAG 649
440 ACCTGCGTGGCGGGAAGCGGTGAACACCGACGTCGAGGAGTCGCGCGCGCTGG 499
650 CTTTAAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAGAGCTTGATTCCTGTGATTG 709
500 ACAAAGTACACCGTCGACATGACCAAGCGCGCGAGGAAGCAAGCTCGATCCGCTGATCG 559
710 GTGCTGATGAGGAGATTAGAAGAGTCTGTGAGGATTCCTTCGAGGAGAACGAACAATC 769
560 GTCCGACGACGAATCCGCGGACCAATCCAGGTCTTCGAGCGCGGACCAAGAACACC 619
770 CTGTGCTTATTTGAGAGCCAGGAGTTGCTGTAATAACAGCTGTGTGAGGTTTGAACAAA 829
620 CGGTGCTGATCGCGGAACCCCGGCTCGGCAAGACCGCATCGTCGAGGCGCTGCGCCAGC 679
830 GGATTGTGAAGAGAGATGTGCCAACAGTCTTACTGATGTGAGATTAAATTTCTGTGACA 889
680 GCATCATCAACGCGGCAAGTGCAGGACCGGCTCAAGGCAAAAGCGCTGCTGGCCCTGGACA 739
890 TGGTGGCTTGTAGTTGCTGGTGTCTAAATACCGAGGAGAGTTTGAAGAAAGGTTGAAATCTG 949
740 TGGGGGCGCTGATCGCGGTGCGCAAGTTCCGCGCGAGTTTCGAGGAACGCTCGAAGCAG 799
950 TTTTGAAGAGATTGAGGACGCTGAAGCGAAAGTGTCTCTTTTATGATGAGATTCAAT 1009
800 TCCTCAACGAACCTGGGCAAGCAGGAAGCGCGGTCTATCTGTTTCATCGACGAACCTGCACA 859
1010 TGGTCTTGTGCTGGGCAAACTGAAGGCTGATGATGATGATGATGATGATGATGATGATGATG 1069
860 CCATGGTGGTGGCGGCAAGCGGCAAGGTTCCATGGAACGCGCGCAACATGCTCAAGCCGG 919
1070 TGTGTAGCTAGAGGCGAGCTGCTGATGCTGCTCAACAGCTTGAAGAAATACAGGAAT 1129
920 CTCTGGCGCGCGGAGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 979
1130 ATGTTGAGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1189
980 ACATCGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1039
1190 GTGTGCTGACACCATTTAGTATCTCTTAGAGGATCTAAGGAGAGATGATGAGGAGATCATG 1249
1040 GCGAGGAAGACACGATCGCATCTCTCGTGGTCTCTAGGAACGCTATGAGTGCACCCAG 1099

QY 1250 GTGTGCGAATCCAGACACAGAGCTCTTATAATGCTGCTCAGCTGTCTGCTGTTACATAA 1309
DB 1100 GGGTGAGCATCACCGACGCGCGATCATCGCGCGGCAAGCTGTGCGACCGCTACATCA 1159
QY 1310 CTGGTCGGCATTTACCGGATAAAGCAATTTGATTTGTTGATGAGGCTTTGTGGAATGTGA 1369
DB 1160 CCGATCGCAACTGCGGACAGGCGCATCGACCTGATCGACGAGCGCGCCAGCCGCTCC 1219
QY 1370 GAGTCAGCTTGATAGTCAACTGAAGAGATTGATAA CTTTGAAGAGAGAGATGCGAGC 1429
DB 1220 GCATGAGATCGACTCCAAGCCAGAGGAACTGGATCGTCTCGACCGTCCGCTGATCCAGC 1279
QY 1430 TCGAAATTTAACTTCAACGCTTGGAAAGGGAAGGATAAGCCAGCAAGCTCCACTTA 1489
DB 1280 TGAAGATCAGCGCGAGCGCTGAAAGAGGAAGACGACGAAGCCACCAAGAGCGCCCTGG 1339
QY 1490 TAGAGTGCAGAAAGAGCTTGATGACCTCGAGAGACAAGCTTCAGCTCTCTCAGATGAAT 1549
DB 1340 CCAAGCTGGAGGAGATATCGTCAAGCTCGAGCGCAATACGCGACCTCGAGGATATCT 1399
QY 1550 ACAGAAAGGAGAAAGAGAGATTGATGATTCGAGGCTTAAACAGAAAGAGAGAGC 1609
DB 1400 GGAAGTCCGAGAAGGCCGAGGTGCAAGGCTCGGCGCAGATCCAGCAGAAGATCGAGCAGG 1459
QY 1610 TCATGTTTTTCTTTCAGAGGCGCAGAAACGAATATGACCTTGCAGAGAGCTGCTGATCTAA 1669
DB 1460 CCAAGCAGAGATGAGGCGCGCGCGCAAGGCGAGCTCGAGAGCATGGCGCGCATCTC 1519
QY 1670 GATATGGCGCAATCAAGAAGTGGAACTGCAATTTGCCCAACTTTGAAGGAACTTCTCTG 1729
DB 1520 AGTACAGACCATCCGACCTGGAAACGACGCTGACAGTGTGACACGACGCGCAAGA 1579
QY 1730 AAGAGATGTGATGCTCAGAGAAAGCTTTGGGCTGGAACATTTGCTGAGGTTGTGAGCC 1789
DB 1580 CCGAGAACCAAGTTGCTGCGCAACCAAGGTGACCGCAGAGGAAATCGCCGAAGTGGTTCCA 1639
QY 1790 GTTGCACAGGAGATCCAGTGCAGAGACTTGGCCAAATGAGAAGGAGAGTTGATTGTC 1849
DB 1640 AGTGAACCGATATCCGTTGCAAGATGCTCGAGGCGAGCGGAGAGCTGCTGCGCA 1699
QY 1850 TTGCTGATAGGTTGATTAAGCGGGTTGTGGGACAGATCAAGCGGTAAATCGAGTTTCTG 1909
DB 1700 TGGAGCAGAGCTGATCGCGAGTGTATCGGCCAGGACGAGCGGTAGTTCGCGGTGCTCA 1759
QY 1910 AGGCAATTTAAGTCAAGGCGCAGACTTGTGTAGGCAACACAGCCAACTCGATCATTTCT 1969
DB 1760 ACGCGTGCCTGCTGCGCGCGCGCTCGCCGATCCGAACCGGCGAGCGGCTCGTTCC 1819
QY 1970 TATTCCTTGGACCAACTGCTGTGGCAAACTGAGCTCGCAAGGCTCTTCTGAGCAGC 2029
DB 1820 TCTTCTCGGCCGACCGGGTGGCAAGACGAGTTGTGCAAGCGCTGCGCGAGTTCC 1879
QY 2030 TGTGATGATGAATAACCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 2089
DB 1880 TCTTCGATACCGAGGAGCGCTGTGTGCGGATCGATATGTCGAGTTTATGAGAGAACT 1939
QY 2090 CTGCTCTCGCTCATTTGGGSCACACAGGCTGATGTTGTTGATGATGATGATGATGATGATG 2149
DB 1940 CCGTGGCGCGCTGATCGCGCGCTCGCGCTACGTCGGCTTTGAGGAAGCGGCTAC 1999
QY 2150 TAACTGAGGCTGTGAGGAGGCGACTTATGCTGTGATGATGATGATGATGATGATGATGATG 2209
DB 2000 TGAACGAGGAGATCGCGCGCAAGCTCTACTCGGTGGTGTGCTGATGATGATGATGATGATG 2059
QY 2210 CTGATGTTGCTGTCTTCAACACTCTGCTCCAAAGTTTTGGATGATGATGATGATGATGATG 2269
DB 2060 CCGATCCGATGATTTCAACATTTCTCTCCAGGTGCTCGAGGACGAGCGCTGACCGACA 2119
QY 2270 GCAAGGAGGAGCAGTCAATTTGAGGAATCGGTGATGATGATGATGATGATGATGATGATG 2329
DB 2120 GTACGGGCGTACGCTGGAATCTCCGCAACACCGTGGTGTGATGATGATGATGATGATGATG 2179

2330 CTGAACACCTCTTGTGAGGCTAACTGGGAAAGTAACTAAATGGAAGTGGCCGGAGCTGTG 2389
 2180 CGGCACAGATCCAGG-----AGCTGCGCGGACCGCGGCGCAACCTGCGCGAG 2230
 2390 TGATGCGGAGGTGAGGAACACTTCAGACCAGAGCTCTTGAAACAGGCTTGAGGAGATTG 2449
 2231 TGATGAGCGCGGTCAATGACACTTCCTGCGGAAATTCATCAACCGGATCGACGAAGTGG 2290
 2450 TGTGTTTTCGACCCCTTTCATGACACAGTTGAGGAAAGTAGCTCGGCTTCAAAATGAAG 2509
 2291 TGTATTCGAGCGCTGCTGCGGACAGATCGCCGCGATCGCGAGATCCAGCTCGGTC 2350
 2510 ACTGTGCTGCGGCTGCTGCGGAGAGGAGTTGCTTTGCGAGTCACTGATGCTGCTTTGG 2569
 2351 GCCTGCGCAAGCCCTGGCGGAGCGGAGCTGAGCCTGCAACTGAGCCAGGAGGCGCTGG 2410
 2570 ACTATATCTGGCAGAGAGTTATGACCCGCTGTATGCTGCTAGGCTTATAGGAGATGA 2629
 2411 ACAAGCTGATCGCGCTGCGCTTCGACCCCGCTCATGGCGCAGCCGCTGAAAGCGGCGCA 2470
 2630 TGGAGAAGAGGTGGTGCACAGAACTGTCAAGAGATGTTGCTGCTGAGGAAATCG 2683
 2471 TCCAGGCTGATCGAGAACCCGCTGGCGCACTGATCTGGCGCGCAATTCG 2524

BULT 13
 09-252-991A-4897/c
 Sequence 4897, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 4897
 LENGTH: 2970
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 09-252-991A-4897

Query Match 17.6%; Score 545.2; DB 4; Length 2970;
 Best Local Similarity 53.7%; Pred. No. 1.4e-147;
 Matches 1157; Conservative 0; Mismatches 988; Indels 9; Gaps 1;
 530 TCAGGATTTGTTGAACGAAGTGGTGTAGCGACGCGGAGGCTAAAGTCTGAGGTTGAGA 589
 2585 TCGGCAAGCTGTGCTCGCGCAGGCGGTCTCGCGCAAGCGCTGAGGATGCGGTGGCCA 2526
 590 AGCTTCGTGGGAAAGAGGGAAGATTGAGAGTCTTCAGGGGACACAAATTTTCAAG 649
 2525 ACCTGCGTGGCGGGAAGCGGTGAACACCGAAGCTGAGGAGTTCGCGCCAGGCGCTGG 2466
 650 CTTTAAAGACTTATGGAAGAGATTGTTGACGACGAGGAGCTTGATCTCTGTGATTG 709
 2465 ACMAGTACCGTGCACATGACCAAGCGCGCGAGGAAGGCAAGCTCGATCCGGTATCG 2406
 710 GTCGTGATGAGGAGATTAGAAGAGTCTGAGGATTTCTTTGAGGAGAAACGAAGAACATC 769
 2405 GTGCGCACGACGAATCCCGCGACCATCCAGGTCTCTGAGCGCGGCGGACCAACAC 2346
 770 CTGTGCTTATGAGAGCGAGGATTGGTAAACAGCTGTGTTGAAAGTTTGAACAAA 829
 2345 CGGTGCTGATCGCGCAACCGCGCTGCGCAAGACCGCCATCTGCGAGGCGCTGGCCAGC 2286
 830 GGATTCGTGAAGAGAGATGCGCAACAGCTTCTACTGATGTGAGATTAATTTCTGTGACA 889

2285 GCATCATCAACGCGCAAGTGCAGCGCGCTCAAGGACAAAGCGCTCTGCGCCCTGGACA 2226
 890 TGGGTGCGTTAGTTGCTGTGCTTAAATACCGAGAGAGATTGTTGAAGAAAGTTGAAATCTG 949
 2225 TGGGGGCGCTGATCGCGGTGCTCAAGTTTCGCGCGGAGTTCGAGGAACGCTGAAGCGAG 2166
 950 TTTTGAAGAAGTTGAGGACGCTCAAGGCAAGTGAATCTCTTTATTGATGAGATTCAAT 1009
 2165 TCCTCAAGCAACTGGCGCAAGCAGGAGCGCGGTATCTCTGTTTCATCGACGACTGACA 2106
 1010 TGGTCTTGTGCTGGCAAACTCAAGGCTGATGATGATGATGATGATGATGATGATGATG 1069
 2105 CCATGCTGCTGCTGCGGCAAGCGGAGGTCCTAGGACGCGCGCAACATGCTCAAGCGG 2046
 1070 TGTGAGTGTAGAGGCGAGCTTCGATGCTCAAGCTGCTCAAGCTGCTCAAGCTGCTCA 1129
 2045 CTCTGCGCGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1986
 1130 ATGTTGAGAAAGATGCTGCTTTCGAGAGAGGTTTCCAAAGTCTATGTTGCGGAGCCAA 1189
 1985 ACATCGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1926
 1190 GTGTGCTGACACCAATGATGATGCTTTCGAGAGGCTCAAGGAGAGATGATGAGGAGAT 1249
 1925 GCGAGGAAGACACGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1866
 1250 GTGTGCGAATCCAAAGACAGAGCTCTTATAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309
 1865 GGGTGAAGTACACGAGCGGCGGATCATCGCGCGGCAAGCTGCTGCGCTGCTGCTGCTGCT 1806
 1310 CTGTGCGCAATTAACCGGATTAAGCAATGATGATGATGATGATGATGATGATGATGATGAT 1369
 1805 CCGATCGCAACTCGCGGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1746
 1370 GAGTCCAGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1429
 1745 GCATGAGAGTGCATCCAGGCGGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1686
 1430 TGGAAATTTGAATTCAGCTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1489
 1585 TGAAGATCGAGCGGCGGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1626
 1490 TAGAGTTCGCGAAAGAGCTTGTGATGCTGAGAGCAAGCTTTCAGCTTCTCAGGATGAA 1549
 1625 CCAAGCTGAGGAGGATATGCTCAAGCTGAGCGGCAATACGCGGCTGCTGAGGATATCT 1566
 1550 ACAGAAAGGAGAGAGAGAGATTCATGATGATTCGAGGCTTAAACAGAGAGAGAGAGAGC 1609
 1565 GGAAGTCCGAGAGGCGGAGGCTGCGGCGGAGATCCAGGAGAGATCCAGGAGGAGGAGG 1506
 1610 TCATGTTTCTTTTTCGAGGAGGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 1669
 1505 CCAGCAGGAGATGAGGCGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1446
 1670 GATAGCGGCAATTCAGAGAGTGAATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1729
 1445 AGTACCAAGACCATCCGAGCCTGGAACCGAGCTGAGATGCTGAGGAGGAGGAGGAGGAG 1386
 1730 AAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
 1385 CCGAGAACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1326
 1790 GTTGAAGAGGATTCAGTGCAGAGACTTGGGCAAAATGAGAGGAGGAGGAGGAGGAGGAGG 1849
 1325 AGTGACCGGTATCCCGGTGCTGAGATGCTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1266
 1850 TTGCTGATGAGGTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
 1265 TGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1206
 1910 AGGCAATTTCAAGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1969
 1205 ACGCCGTGCGCGCTTTCGCGCGCGGCTGCGCGATCCGAACCGGCGGAGCGGCTGCTTCC 1146

```

1970 TATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCCAGGCTCTTGTGAGCAGC 2029
1145 TCTTCTCGGCCCGGACCGGGTGGGCAAGCCGAGTGTGCAAGGCGCTGGCGGAGTTC 1086
2030 TGTTCATGATGAAACCTCTTAGTTCGGATTCATATGTCGGAATATATGGAACAACT 2089
1085 TCTTCATACCGAGGAGCGCTGGTGGGATCGATATGTCGAGTTCATGGAAGAACT 1026
2090 CTGCTCTCGCTCATTTGGGGACCAACCGAGGTATGTTGGTCAAGGAGGTGGACAC 2149
1025 CGGTGCGCGCTGATCGGCGCGCTCCGGGCTACGTCGGCTTTGAGGAAGCGGCTACC 966
2150 TAACTGAGCTGTGAGGAGCGACCTTATGTGTATCTTGTGATGAGTGGAGAAG 2209
965 TGACCGAGGCGATCCCGCGCAAGCCCTACTCGGTGGTGTGTCGACGAGGTGGAGAAG 906
2210 CTATGTTGTCTCTTCAACACTCTCTCCAGTCTTGGATGATGTCGATTCGACAGAG 2269
905 CCATCCGGATGATTCACATCTCTCCAGTGTCTGAGGACGACGCTGACCGACA 846
2270 GGCAAGCGAGGACGATGCAATTCAGAACTCGGTGATATCATGATCAAACTTTGGTG 2329
845 GTCAAGCGGCTACGTTGGACTTCGCAACACCGTGGTGGTGTGATGACCTCCAACTCGGTT 786
2330 CTGAACACCTCTTGCAGGCGCTAACTGGGAAAGTAACTGAATGAGTGGCCCGGACTGTG 2389
785 CGGCACAGATCCAGG-----AGCTGGCGGGGACCGCGAGCGCAACGTGCGGAG 735
2390 TGATCGGAGGTGAGGAAACACTTCAGACACAGAGCTCTTGAACAGGCTTGACGAGATTG 2449
734 TGATGACGCGTCAATGACACTTCGCTCGGAAATTCATCAACGGATCGACGAGTGG 675
2450 TGGTGTTCAGCCCTTTCACATGATGACGATTCAGGAAAGTACGTCGGCTTCAAAATGAAG 2509
674 TGGTATTCAGCGCGTGGTTCGCGAGCAGATCGCGGATCGCGGAGTCCAGCTCGGTC 615
2510 AGTTGCTGTCGCGCTTGTGAAAGAGGAGTGTCTTTCGAGCTCACTGATGCTGTTGG 2569
614 GCTGGCGAGCGCTTGGCGAGCGGAGCTGAGCCTGGAATGAGCCGAGGAGCGCTGG 555
2570 ACTATATCTTGGCAGAGATATGACCCGCTGTATGCTGCTAGGCTATAGGAGATGGA 2629
554 ACAAGCTGATCGCGTTCGCTTCGACCGGCTATGCGCGACGCGCGCTGAGCGGGCA 495
2630 TGGAGAGAGGTGGTGAAGAACTGTCAAAGATGTTGTGCGTGAGGAATTCG 2683
494 TCCAGCGCTGGATCGAAGACCGGCTGGCGCAACTGATCTGCGCGGCAAAATTCG 441

```

LT 14

8-249-380-1
 Sequence 1, Application US/08249380
 Patent No. 5827685
 GENERAL INFORMATION:
 APPLICANT: Lindquist, Susan
 TITLE OF INVENTION: Methods and Compositions of Genetic
 TITLE OF INVENTION: Stress Response Systems
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P. O. Box 4433
 CITY: Houston
 STATE: TX
 COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/249,380

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 07/710,187
; FILING DATE: 31-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3727 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-249-380-1

Query Match      16.0%; Score 495.4; DB 1; Length 3727;
Best Local Similarity 52.2%; Pred. No. 4.3e-133;
Matches 1351; Conservative 0; Mismatches 1176; Indels 60; Gaps 9;

QY 155 AATCGAAGATGAATCCAGAGAAATTCACACACAGACAAACGAGACAAATTCGTACAGCTC 214
DB 996 AATATATGAACGACCAACGCAATTTACAGAAAGGCTCTAACGATTTTGACGTTGGCTC 1055
QY 215 ATGAGCTAGCTGTAATGAGGACATCTCAATTCATCTCTTTGCAATTTAGCTTGGTCTT 274
DB 1056 AAAAATTTGGCTTCGGATCATACTCAATTCACAAATTAACCTATACATATTTCTAGCTGCCT 1115
QY 275 TGATCTCTGATCCACCGGTATAT-----TTCTCTCAAGCAATCTCTAGTGGCGGTGGCG 328
DB 1116 TCATTGAACGCGCAGAGATGGATCAGTCCCTTACCTACAGAACTCTAATTTGAGAGGGCC 1175
QY 329 AGAAGCAGCTCAATCTGCTGAAGAGTATCAATCAAGCCTTGAAGAAAGCTTCTTTCAC 388
DB 1176 GTTACGACTATGATCTTTTCAAGAAAGTGGTAAATAGAAATCTAGTAAAGAAATTCCTCAAC 1235
QY 389 AATCTCTCTCACCTGATGATATTCAGCGAGTCTAGTCTTATTAAGGTCAATTCGTCGTG 448
DB 1236 AGCACTGACCTTCGGAGATACTCCAAGTATGCTTTGGGGAAGTCCCTTCAAGAGC 1295
QY 449 CTCAGCTGCTCAGAGTCAAGAGTGTGATCTCATTTGGCTGTTGACCACTTGATTAAG 508
DB 1296 CTGCTAAGATTCAAAACCAACAGAGGACTCATTTATAGCGCAAGACCATATATTTGTTG 1355
QY 509 GTCTTCTTGAAGATCTCAATCAGGATTTGTTGAACGAACTCGGTGTAGCGACGCGGA 568
DB 1356 CTCATTCAATGATTCGTCTATTTCAGCAATAATTTAAGGAAGCTCAAGTAGATATTGAG 1415
QY 569 GGGTAAAGTCTGAGGTTGAGAACCTTCGCGGAAAGAGGAGAAAGTTGAGAGTGTCT 628
DB 1416 CCATCAGCAACGAGCTCTTGAACCTTCGCTGTAACACTAGAAATTTGCTCTGCTGGCGCTG 1475
QY 629 CAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGC-- 686
DB 1476 ATACGACACACCTTTGGATATTTTCAAGTACGCCATTTGATATGATGATGACGTGAGCAGCTC 1535
QY 687 ----AGGGAAGCTTGATTCCTGATTTGGTGTGATGAGGAGATTTAGAAAGTGTGAGGA 742
DB 1536 GTCAGGTAATCTGACCCCTGTCTATCGCGCGTGAAGAAAGAAATTAAGAGCACTATTAGAG 1595
QY 743 TTCTTTTCGAGGAGAACGAAAGCAATCTCTGCTTATTCGAGAGCCAGAGTTGGTAAAG 802
DB 1596 TTTTAGCAAGAGAAATTAAGTCCAACTGTTTAAATTTGAGCAGGATCGGTAAAG 1655
QY 803 CAGCTGTGGTTGAAGGTTTAGCAAAAGGATTTGTAAGAGGATGTGAAAGAGATGTGCCCAACAGTCTTA 862
DB 1656 CCGCTATTATGAAGGTTGCTCAAGAAATCAATGACGATGACGTTCCCATTTCTTAC 1715

```

863 CTGATGTCAGATTAATTTTCGTTGGACATGGTGGCTTAGTTGCTGGTCTAAATACCGAG 922
1716 RAAGCGCTAAATTTTCAGTCTAGATTTGGCGCATTAACCGCAGGTCTAAATACAAAG 1775
923 GAGAGTTTGAAGAAAGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGAAGGCAAG 982
1776 GTGATTTGGAAGAAAGATTCGAAGGTTGTTTGAAGGAAATCGAAGAAATCAAGACTCTAA 1835
983 TGATTTCTCTTTATTCATGAGATTCATTTGTTGTTCTGTTGCTGGCAAAACTGAAGGTCGA 1042
1836 TTGTTGTTATTCATTCATGAGAAATTCATGTTTAATGCGTAATGGTAA-----GG 1883
1043 TGGATGACGCTAATCTGTTCAAGCCCATTTAGCTAGAGGCGAGCTTCGATGCAATTTGGTG 1102
1884 ACGACGCTGCTAAACATCTTGAAGCGCAGCTTTGTCAGAGGCAATTTGAAGGCTCATCGGTG 1943
1103 CTACACGCTTGAAGAAATACAGGAATATGTTGAGAAAGATGCTGCCCTTTGAGAGGAGGT 1162
1944 CACCACCAATTAACGAATATAGATCTATTTGTTGGAAGAGGATGGTGGCTTTGAAAGAAGAT 2003
1163 TCCAAACAAGTCTATGTTGCGGAGCCAAAGTGCCTTGACACCATTTAGTATCTTTAGAGGAC 1222
2004 TCCAGAAATTTGAAGTCTGCTGAACTGAGTGTGAGACAAACAGTGGCCATATTTAGAGGTC 2063
1223 TCAAGGAGAGTATGAGGACATCATGTTGTCGGAATCCAAAGACAGAGCTCTTTATAAATG 1282
2064 TGCAACCAAGATGATGAATATACATCATGTTGTTAAGGATTTCTGATAGCGCTTAGTCACTG 2123
1283 CTGCTCAGCTGCTGCTGTTACATAAATGTTGCTGCGCATTTACCGGATTAAGCAATTTGAT 1342
2124 CTGCTCAATTTAGCCAAAGCTTTACTTGGCATATGAGAAATTTCCAGATTTCTGTTGGATT 2183
1343 TGGTTGATGAGGCTTTGCGAATGTGAGAGTCCAGCTTTGATAGTCAACCTGAAAGAGATTG 1402
2184 TAGTTGATATTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2243
1403 ATACCTTTGAA-AGGAAGAGGATGAGCTGGAATTTGAATTTCAAGCTTTGGAAGGAG 1461
2244 ATTCGAAGGACAGTCAATTTGATTTGATTTCAAGTATGAGATAAAGCTTTAGAGAGAGAG 2303
1462 AAGGATAAAGCCAGCAAGCTCGACTTATAGAGTGGCGGAAG-----AGCTTATGACC 1516
2304 TAGAGTGGGACTCCACACTAAAGAAAGTTAAAGTTAGCTAGACAGAGGAAGCTTCAT 2363
1517 TGAGAGACAAAGCTTCAGGCTCTCAGATGAATACAGAAAGGAGAGAGAGATTTGATG 1576
2364 TGCAAGAAAGATTTGAACCTCTTAAGCAACGTTTAAATGAAGAAAGATGCGCATGAG 2423
1577 AGATTCGAAGGCTTAAACAGAAAGAGAGAGCTCATGTTTCTTTGAGAGGAGGAGAC 1636
2424 AATTGACACAAGCTTAAAGAAAGATTTGGATGAATCTGGAAGAAAGAGGCTTTAGCTGAAC 2483
1637 GAAGATATGACCTTCGAAGAGCTGCTGATCTAAGATATGCGCAATTCAGAAAGTGAAT 1696
2484 CTAGATATGATCTGCTGACCGCTGCTGATTTAAGGTACTTCCCATCCAGATCAAA 2543
1697 CTGCAATTCCTCACTTCAAGAACTCTTCTGAGAG-----AATG 1738
2544 AGCAAAATCGAAAGCTTCAAGATCAGTTGCTGAGGAAGAGAGAGCTGCTGGTCCCACT 2603
1739 TGATGCTCAGAGAAAGCTTGGGCTGAAACATTTGCTGAGGTTGTGAGCGCTTGGACAG 1798
2604 CCATGATCCAAATGTTGCTGATTCAGACACCATTTCTTGAAGAGCTGCAAGATTTGACTG 2663
1799 GGATTTCCAGTACAGAGCTTGGCCAAATGAGAGGAGAGGTTGATTTGCTGCTGATA 1858
2664 GTATCCCTGTTAAGAGTTTGCAGAACTGGAATGAAATGAAATTTGATTCATATGGAACGTG 2723
1859 GGTTCGATAAGCGGTTGTGGACAGAAATCAAGCGGTAATTCGAGTTTCTGAGGCAATTC 1918
2724 ACTTATCATCTCAAGTCTGGGCGCAATTTGGATGCCATTTAAAGCTTTTCCCAATGCCGTTA 2783
1919 TAAGGTCAGGCGAGGACTTGGTAGGCGACAAACAGCCAACTGGATCATCTTATCTCTTG 1978

2784 GATTTGCTAGATCAGGTTTAGCTTAATCCAGGCAACAGC---ATCCTTCTTATTTTAG 2840
1979 GACCAACTGTTGTCGCAAAATGAGCTCGCAAGGCTCTTCTGAGCAGCTGTTTGTATG 2038
2841 GTTTGTCGGCTTCGGCTTAAATCGAATTTGGCTTAAAGGTTGCTGATTTTGTAAATG 2900
2039 ATCAAAACCTCTTAGTTTCGGATGATATGTCGGAATATATGNAACAACACTCTGCTCTTC 2098
2901 ATGAGGACATGATGATCAGGCTCGATTTGTTCTGAATTAAGCGAAGATATCGGCTCTCTA 2960
2099 GCCTCATTTGGGGCACCACCAGGCTATGTTGGTCACGAGGAAGGTGACAACTAACTGAGG 2158
2961 AGTTGTTGGGTACCAACGCGAGGTTATGTCGGGTACGATGAAGGTGCTTTTAACTAACCC 3020
2159 CTGTGAGGAGCGACCTTATTTGCTCACTCTTCTGATCAAGTGAAGGAGGCTCATGTTG 2218
3021 ARCTGCAATCAAAACCATATCTCGGTTTGTGTTATTCGATGAAGTAGAAAGGACATCCTG 3080
2219 CTGCTTTCAACACTCTGCTCCAAAGTTTGGATGATGCTGATGATGACAGACGGCAAGGCA 2278
3081 ATGTTTGTGCTGCTGCTACAAATGTTGATGACGCTAGAAATTTCTTCTGCTCAAGGTA 3140
2279 GGCAGTTCGATTTTCAAGAACTCGGTGATATCATGACATCAAACTTGGTGTGCTGAACACC 2338
3141 AGACGATCGACTGTTCCAAATTTGTTGTCATCATGACTTCCAATCTAGGTGCTGAATTTA 3200
2339 TCCTTCGAGGCTTAACCTGGGAAGTAAACATGGAAGTGGCCCGGACTGCTGATGCGGG 2398
3201 TCAATTC---TCAACAAGGATCAAGATCCAAAGATTTTACCAGAAATTTGGTCAATGGG 3257
2399 AGGTGAGGAAACACTTTTCAGACAGAGCTCTTGAACAGGCTTTCAGAGATTTGCTGTTG 2458
3258 CTGTAGGCAACATTTTCAGACAGATTTTGAACAGATTTCTAGTATAGTCAATTTCA 3317
2459 ACCCTTTTCACATGACCCAGTTGAGGAAAGTAGCTGGCTTCAAAATGAAGACGTTGCTG 2518
3318 ACAAGCTATCTAGAAAGCTATTTCAAGATGCTGATATTCGTTTGAAGGAAATTTGAAG 3377
2519 -----TCCGGCTTCTGGAAGAGGAGTTGCTTTGGCAGTCACTGATGCTGCTTGGACT 2572
3378 AGAGATTCAGGCAAAATGATAAATTAACATTAAGATTTGAAATTTAACTCAAGAGGCCAAG 3437
2573 ATATCTTGGCAGAGGTTATGACCCGCTGATGCTGCTAGGCTTATAGGAGATGATGG 2632
3438 TCTTGGCCAAATATGTTATTCGATGATATGCTGACGCTCCACTGACAGGTTAATTC 3497
2633 AGAAGAGGTTGTCAGAACTGTCGAAGATGTTGCTGAGGAAATTCGATGAAACT 2692
3498 AAAACGAAATTTTGAACAAACTGGCACTAAGGATCTTTAAAGAAATGAATCAAGGATAAG 3557
2693 CCACTGT 2699
3558 AACTGT 3564

RESULT 15
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2

Search completed: February 12, 2004, 22:31:14
Job time : 236 secs

Search completed: February 12, 2004, 22:31:14
Job time : 236 secs

1	2732.8	88.0	2736	10	US-09-938-942A-235	Sequence 235, App
2	2732.8	88.0	2736	12	US-09-938-942A-235	Sequence 235, App
3	807	26.0	3942	12	US-10-310-154-113	Sequence 113, App
4	672.8	21.7	6491	10	US-09-070-927A-156	Sequence 156, App
5	668.2	21.5	43980	12	US-10-398-221-5	Sequence 5, Appli
6	668.2	21.5	3011208	12	US-10-398-221-2058	Sequence 2058, Ap
7	662	21.3	2586	12	US-10-369-493-43449	Sequence 43449, A
8	659.8	21.2	2604	12	US-10-369-493-42135	Sequence 42135, A
9	651.8	21.0	3000	12	US-10-310-154-117	Sequence 117, App
10	651.4	21.0	2586	12	US-10-369-493-41201	Sequence 41201, A
11	650.6	21.0	2571	12	US-10-369-493-32853	Sequence 32853, Ap
12	649.4	20.9	2571	9	US-09-815-242-7037	Sequence 7037, Ap
13	649.4	20.9	1830121	12	US-10-329-670-1	Sequence 1, Appli
14	649.4	20.9	1830121	15	US-10-329-960-1	Sequence 1, Appli
15	649	20.9	2571	12	US-10-369-493-33061	Sequence 33061, A

Qy	1363	AATGTGAGAGTCCAGCTTGTAGTGTCAACTCTGAAGAGATTGATAACTTTGAAGGAGAGAGG	1422
Db	1201	AATGTGAGAGTCCAGCTTGTAGTGTCAACCTTGAAGAGATTGATAACTTTGAAGGAGAGAGG	1360
Qy	1423	ATGCAGCTGGAATTTGAACCTTCACGCTTGGAAAGGAGAAAGATAAAGCCAGCAAGACT	1482
Db	1261	ATGCAGCTGGAATTTGAACTTTCACGCTTGGAAAGGAGAAAGATAAAGCCAGCAAGACT	1320
Qy	1483	CGACTTATAGAGTGGGAAAGAGCTTTGATGACCTGAGAGACAAGCTTACGCTTCTCAAG	1542
Db	1321	CGACTTATAGAGTGGGAAAGAGCTTTGATGACCTGAGAGACAAGCTTACGCTTCTCAAG	1380
Qy	1543	ATGAATAACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGGCTTAAACAGAAAAAG	1602
Db	1381	ATGGAATAACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGGCTTAAACAGAAAAAG	1440
Qy	1603	GAAGAGCTCATGTTTCTTTGTCAGGAGGCGAGAAACGAAGATATGACCTTGCAGAGCTGCT	1662
Db	1441	GAAGAGCTCATGTTTCTTTGTCAGGAGGCGAGAAACGAAGATATGACCTTGCAGAGCTGCT	1500
Qy	1663	GATCTAAGATATGGCCCAATTCGAAGAGTGGAAATCTGCAATTCGCCCAACTTGAAGAACT	1722
Db	1501	GATCTAAGATATGGCCCAATTCGAAGAGTGGAAATCTGCAATTCGCCCAACTTGAAGAACT	1560
Qy	1723	TCTTCTGAAGAGAAATGTGATGCTTCACAGAAAAAGTTTGGGCTTGAAACACATTTGCTGAGTT	1782
Db	1561	TCCTTCTGAAGAGAAATGTGATGCTTCACAGAAAAAGTTTGGGCTTGAAACACATTTGCTGAGTT	1620
Qy	1783	GTGAGCGGTGGACAGGAAATCCAGTGAACAGACTTTGGCCCAAAATGAGAAAGAGAGTTG	1842
Db	1621	GTGAGCGGTGGACAGGAAATCCAGTGAACAGACTTTGGCCCAAAATGAGAAAGAGAGTTG	1680
Qy	1843	ATTGGTCTTGCTGATAGGTTGCATAAGACGGGTTGTGGGACAGAAATCAAGCGGTAAATGCA	1902
Db	1681	ATTGGTCTTGCTGATAGGTTGCATAAGACGGGTTGTGGGACAGAAATCAAGCGGTAAATGCA	1740
Qy	1903	GTTTCTGAGGCAATTTCTAAGGTCAAGGCGAGACTTTGGTAGGGCAACAACGCCAACTGGA	1962
Db	1741	GTTTCTGAGGCAATTTCTAAGGTCAAGGCGAGACTTTGGTAGGGCAACAACGCCAACTGGA	1800
Qy	1963	TCAATCTTAATTCCTTGGACCAACTGGTGTGTGCAAACTGAGCTGCGCAAGGCTCTTGCT	2022
Db	1801	TCATCTTAATTCCTTGGACCAACTGGTGTGTGCAAACTGAGCTGCGCAAGGCTCTTGCT	1860
Qy	2023	GAGCAGCTGTTTGATGATGAAACCTCTTAGTTTCGGAATTGATATGTTCGGAATATATGGA	2082
Db	1861	GAGCAGCTGTTTGATGATGAAACCTCTTAGTTTCGGAATTGATATGTTCGGAATATATGGA	1920
Qy	2083	CAACACTGTCTCTCGCCTCATTTGGGCAACACACAGGCTATGTGGTCAACGAGAGGT	2142
Db	1921	CAACACTGTCTCTCGCCTCATTTGGGCAACACACAGGCTATGTGGTCAACGAGAGAGGT	1980
Qy	2143	GGACAACTAACTGAGGCTGTGAGGAGCGCACTTATTTGTGTCACTCTTTGATGAAGTG	2202
Db	1981	GGACAACTAACTGAGGCTGTGAGGAGCGCACTTATTTGTGTCACTCTTTGATGAAGTG	2040
Qy	2203	GAGAAGGCTCATGTTGCTGCTCTTCAACACTCTGTCTCCAAAGTTTGGATGATGGTCAATTG	2262
Db	2041	GAGAAGGCTCATGTTGCTGCTCTTCAACACTCTGTCTCCAAAGTTTGGATGATGGTCAATTG	2100
Qy	2263	ACAGAGGCGCAAGGAGGACAGTTCGATTTTCAGAACTCGGTGATATCATGACATCAAC	2322
Db	2101	ACAGAGGCGCAAGGAGGACAGTTCGATTTTCAGAACTCGGTGATATCATGACATCAAC	2160
Qy	2323	CTTGGTGTGAAACACTCTCTTCAGGGCTTAATCTGGGAAAGTAAACATGGAAGTGCCCCG	2382
Db	2161	CTTGGTGTGAAACACTCTCTTCAGGGCTTAATCTGGGAAAGTAAACATGGAAGTGCCCCG	2220
Qy	2383	GACTGTGTGATGGGGAGGTGAGGAAACATTTTCAGCCAGAGCTCTTTCGACAGGCTTGAC	2442
Db	2221	GACTGTGTGATGGGGAGGTGAGGAAACATTTTCAGCCAGAGCTCTTTCGACAGGCTTGAC	2280
Qy	2443	GAGATTTGGTGTTCGACCCCTTTTCATATGACCAGTTTGAGGAAAGTAGCTCGGCTTCAA	2502

1423 ATGAGCTGGAAATTGAACCTTCAGCCCTTGGAAAGGAGAGATGAAGCCAGCAAGCT 1482
 1261 ATGAGCTGGAAATTGAACCTTCAGCCCTTGGAAAGGAGAGATGAAGCCAGCAAGCT 1320
 1483 CGACTTATAGAGGTGCGGAAAGAGCTTTGATGACCTGAGAGACAGCTTTCAAGCTTCAAG 1542
 1321 CGACTTATAGAGGTGCGGAAAGAGCTTTGATGACCTGAGAGACAGCTTTCAAGCTTCAAG 1380
 1543 ATGAAATACAGAAAGGAGAGAGAGATTTGATGAGATTCGAGGCTTTAAACAGAAAGA 1602
 1381 ATGAAATACAGAAAGGAGAGAGAGATTTGATGAGATTCGAGGCTTTAAACAGAAAGA 1440
 1603 GAAGAGCTCATGTTTTCTTTGACAGAGCGAGCAAGCAAGATATGACCTTTGCAAGAGCTGCT 1662
 1441 GAAGAGCTCATGTTTTCTTTGACAGAGCGAGCAAGCAAGATATGACCTTTGCAAGAGCTGCT 1500
 1663 GATCTAAGATATGCGCAATTCAGAGAGTGAATCTGCAATTTGCCCACTTTGAAGGAACT 1722
 1501 GATCTAAGATATGCGCAATTCAGAGAGTGAATCTGCAATTTGCCCACTTTGAAGGAACT 1560
 1723 TCTTCTGAGAGATGCTGCTCAGAGAGAGCTTGGCCCTGACACATTTCTGAGGTT 1782
 1561 TCTTCTGAGAGATGCTGCTCAGAGAGAGCTTGGCCCTGACACATTTCTGAGGTT 1620
 1783 GTGAGCGCTTGGACAGAGATTCAGTGAAGAGAGCTTGGCCCAATGAGAGGAGAGGTTG 1842
 1621 GTGAGCGCTTGGACAGAGATTCAGTGAAGAGAGCTTGGCCCAATGAGAGGAGAGGTTG 1680
 1843 ATTTGGTCTGATAGTGTGATGAAGCGGTTGTTGGGACAGAAATCAAGCGGTAATGCA 1902
 1681 ATTTGGTCTGATAGTGTGATGAAGCGGTTGTTGGGACAGAAATCAAGCGGTAATGCA 1740
 1903 GTTTCTGAGGCAATTCAGAGTCAAGGCGAGAGCTTGGAGGCGACCAACCAAGCTTGA 1962
 1741 GTTTCTGAGGCAATTCAGAGTCAAGGCGAGAGCTTGGAGGCGACCAACCAAGCTTGA 1800
 1963 TCAATCTTATTCCTTGGACCAACTGGTGTGGGCAAACTGAGCTTCGCAAGGCTTCTGCT 2022
 1801 TCAATCTTATTCCTTGGACCAACTGGTGTGGGCAAACTGAGCTTCGCAAGGCTTCTGCT 1860
 2023 GAGAGCTGTTGATGATGAAGAACTCTTAGTTCGATGATGATGTCGGAATATATGGA 2082
 1861 GAGAGCTGTTGATGATGAAGAACTCTTAGTTCGATGATGATGTCGGAATATATGGA 1920
 2083 CAACACTCTGCTCTCGGCTCATTTGGGCGACACACAGGATGTTGTCAGAGGAGGT 2142
 1921 CAACACTCTGCTCTCGGCTCATTTGGGCGACACACAGGATGTTGTCAGAGGAGGT 1980
 2143 GGACAACTAACTGAGGCTGTGAGGAGGCGACCTTATTTGTGTCATACCTTTTGTGAAGTG 2202
 1981 GGACAACTAACTGAGGCTGTGAGGAGGCGACCTTATTTGTGTCATACCTTTTGTGAAGTG 2040
 2203 GAGAGGCTCATGTTGCTCTTCAACTCTGCTCCAGTTTGGATGATGTCGATG 2262
 2041 GAGAGGCTCATGTTGCTCTTCAACTCTGCTCCAGTTTGGATGATGTCGATG 2100
 2263 ACAGAGGCGCAAGGAGGAGAGAGTGTGATTTGAGGAGCTCGGTGATAATCATGACATCAAC 2322
 2101 ACAGAGGCGCAAGGAGGAGAGAGTGTGATTTGAGGAGCTCGGTGATAATCATGACATCAAC 2160
 2323 CTTGCTGTGACACCTCTTTGCGAGGCTAACTGGGAAAGTAACTGGAAGTGGCCCGG 2382
 2161 CTTGCTGTGACACCTCTTTGCGAGGCTAACTGGGAAAGTAACTGGAAGTGGCCCGG 2220
 2383 GACTGTGTGATGCGGAGGTGAGGAACTCACTCAGACAGAGCTTTGACAGGCTTGAC 2442
 2221 GACTGTGTGATGCGGAGGTGAGGAACTCACTCAGACAGAGCTTTGACAGGCTTGAC 2280
 2443 GAGATTTGTTGTTGCTGACCCCTTTTCACTGACAGCTTGGAGAAAGTGTGCGCTTCAA 2502
 2281 GAGATTTGTTGTTGCTGACCCCTTTTCACTGACAGCTTGGAGAAAGTGTGCGCTTCAA 2340

QY 2503 ATGAAAGAGCTTGTGCTCGGCTTCTGCTGAAAGAGAGGTTGCTTTGGCAGTCACTGATGCT 2562
 Db 2341 ATGAAAGAGCTTGTGCTCGGCTTCTGCTGAAAGAGAGGTTGCTTTGGCAGTCACTGATGCT 2400
 QY 2563 GCTTTGGACTATATCTTTGGCAGAGATTATGACCCGGTGTATGTTGGTGTAGGCTTATAAGG 2622
 Db 2401 GCTTTGGACTATATCTTTGGCAGAGATTATGACCCGGTGTATGTTGGTGTAGGCTTATAAGG 2460
 QY 2623 AGATGGATGGAGAAAGGTTGGTGAACAGAACTGTCAAAGATGGTTGCTGCGTGAGGAATC 2682
 Db 2461 AGATGGATGGAGAAAGGTTGGTGAACAGAACTGTCAAAGATGGTTGCTGCGTGAGGAATC 2520
 QY 2683 GATGAAACTCTCACCTGTTTACATAGATGAGCGGCTGGTGTCTTGTGTACCGGAGTAA 2742
 Db 2521 GATGAAACTCTCACCTGTTTACATAGATGAGCGGCTGGTGTCTTGTGTACCGGAGTAA 2580
 QY 2743 AGTGAAGTCTTAGTGGAGCGCTTCAACAGGCAAGAGTCAAGATGCTGATTCATATTGCT 2802
 Db 2581 AGTGAAGTCTTAGTGGAGCGCTTCAACAGGCAAGAGTCAAGATGCTGATTCATATTGCT 2640
 QY 2803 AACGGCCCAAGAGAGAGTGTGACGCTCAGCGGCTGAAGAGATGAGGATCGAGGAATA 2862
 Db 2641 AACGGCCCAAGAGAGAGTGTGACGCTCAGCGGCTGAAGAGATGAGGATCGAGGAATA 2700
 QY 2863 GAAGATGACGATAATGAGGAATGATCGAGGATTA 2898
 Db 2701 GAAGATGACGATAATGAGGAATGATCGAGGATTA 2736

RESULT 3
 US-10-310-154-113
 ; Sequence 113, Application US/10310154
 ; Publication No. US20030233670A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; APPLICANT: Chomet, Paul S.
 ; APPLICANT: Adams, Thomas H
 ; APPLICANT: Ruff, Thomas G.
 ; APPLICANT: Agarwal, Ameeta K.
 ; APPLICANT: Ahrens, Jeffrey B.
 ; APPLICANT: Ball, James A.
 ; APPLICANT: Banu, G.
 ; APPLICANT: Bell, Erin
 ; APPLICANT: Boddupalli, Raghava
 ; APPLICANT: Deikman, Jill
 ; APPLICANT: Deng, Molian
 ; APPLICANT: Dong, Jinzhao
 ; APPLICANT: Duff, Stephen M.
 ; APPLICANT: Galligan, Meghan M.
 ; APPLICANT: Hinchey, Brenda S.
 ; APPLICANT: Huang, Shihshieh
 ; APPLICANT: Johnson, G. Richard
 ; APPLICANT: Jung, Vincent
 ; APPLICANT: Kretzmer, Keith A
 ; APPLICANT: Laccetti, Lucille B.
 ; APPLICANT: Lai, Chao-Qiang
 ; APPLICANT: Lee, Gary
 ; APPLICANT: Lin, Jie-Yi
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Lu, Bin
 ; APPLICANT: Luethy, Michael M.
 ; APPLICANT: Lund, Adrian
 ; APPLICANT: Madson, Linda L.
 ; APPLICANT: Malloy, Kathleen A.
 ; APPLICANT: McKiel, Christine L.
 ; APPLICANT: Miller, Philip W.
 ; APPLICANT: Padnavathi, Manchikanti
 ; APPLICANT: Parnell, Laurence D.
 ; APPLICANT: Start, William G.
 ; APPLICANT: Tennesen, Dan
 ; APPLICANT: Vidya, K.R.
 ; APPLICANT: Wang, Haiyun
 ; APPLICANT: Xin, Zhanguo

APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping
 APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li
 TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
 FILE REFERENCE: 38-15(52796)B
 CURRENT APPLICATION NUMBER: US/10/310,154
 PRIOR FILING DATE: 2002-12-04
 PRIOR APPLICATION NUMBER: 60/337,358
 PRIOR FILING DATE: 2001-12-04
 NUMBER OF SEQ ID NOS: 736
 Q ID NO 113
 LENGTH: 3942
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (436)..(1773), (1878)..(2159), (2281)..(2711)..(2990), (3079)..(3276),
 OTHER INFORMATION: unsure at all n locations
 NAME/KEY: unsure
 LOCATION: (1)..(3942)
 OTHER INFORMATION: unsure at all n locations
 0-310-154-113
 Every Match 26.0%; Score 807; DB 12; Length 3942;
 Local Similarity 58.5%; Pred. No. 28-217;
 Matches 1918; Conservative 0; Mismatches 845; Indels 514; Gaps 9;

990 CCGCCGAGGAGATCGCCGCGTCTGTCGCAATCTCTCGCGCGCACCAAGATAACCC 1049
 771 TGTCTTATTTGGAGAGCCAGAGTTGGTAAACAGCTGTGTGGTTGAAGTTTAGCAAAAG 830
 1050 CGTCTCATCGCGAGCCCGCGCGTTGCAAGACGGCGCTGTCGAGGCGCTCGCGCAGCG 1109
 831 GATTGTCAAAAGGAGATGTCGCCCAACAGTCTTACTGATGTGAGATTATTTCTGTTGGACAT 890
 1110 CATGTTTCGGCGGAGAGTGCCTCCAGTAACCTCTCGAGCTCGGCTCATCGCTCGACAT 1169
 891 GGGTGCCTTATGTTCTGCTGCTAAATACCGAGGAGAGTTTGAAGAAAGTTGAATCTGT 950
 1170 GGGCGCTCTGTCGCGCGCGCAAGTACCGCGCGAGTTCGAGGAGCGCTCAAGGCGGT 1229
 951 TTTGAAGAAGTTGAGGAGCGCTGAGSCAAAGTATCTCTTTTATGATGAGATTCAATTT 1010
 1230 GCTCAAGGAGGTGGAAGAGCGCGAGGGAAGGTCAATCTCTTCATCGACGAGATACACCT 1289
 1011 GGTCTTGTGCTGGCCAAACTCAAGGTCGATGGATGACAGTAACTCTGTTCAAGGCCAT 1070
 1290 CGTCTGGCGCGGCGAGGAGCGAGGTTTCATGACGCGGCCACCTGTTCAAGCCAAT 1349
 1071 GTTAGCTAGAGGCGAGCTTCGATGTCATTGCTCAACAGCTTTGAAGAAATACAGAAATA 1130
 1350 GCTGGCGAGGCGAGCTCAGGTGTCATCGCGCGCACCAAGCTGAGGAGTACCGCAAGTA 1409
 1131 TGTGGAAGAAGTGTGCTTTCGAGGAGGTTTCCAAAGTCTATGTTGGCGAGCGAAG 1190
 1410 CGTGGAAGAAGCGAGCGTTCGAGCGCGGTTCCAGCAGGTGTTCTGTCGCGAGCGGAG 1469
 1191 TGTGCTGACACCATAGTATCTTGAAGGACTCAAGGAGAGTATGAGGGAATCATGCG 1250
 1470 CGTGGCGGACACCGTCAGCAATCTGAGGGGCTCTAGAGAGAGTACGAGGGGACCATGG 1529
 1251 TGTGCGAATCCAAAGACAGAGCTCTTATAAATGCTGCTCAGCTCTGCTGCTGTTACATAAC 1310
 1530 CGTGAGGATCCAGGACCGCGCTCGTGTGTCGCGGACAGCTCTCGCGAGGTTACATCAT 1589
 1311 TGTGCGCATTTTACCGGATAAAGCAATGATTTGTTGATGAGGCTTGTGCGAATGTGAG 1370
 1590 GGGTTCGCGACTGCTGCTGCAAGCCATAGACTTGTGGAAGGAGGCTTGTGCGCAATGTGAG 1649
 1371 AGTCCAGCTTGTAGTCAACCTCAAGAGATTGATAACCTTGAAGGAGAGGATGCGAGCT 1430
 1650 GGTGCGAGCTCGACAGCGCGGAGAGATGATGATGATGATGATGATGATGATGATGAT 1709
 1431 GGAATTTGAATTTCAAGCTTGAAGGAGGAGAGGATTAAGCCAGCAAGAGCTGCACTTAT 1490
 1710 TGAGGTTGAGCTCCACGCGCTCGAAGGAGAGGAGCAAGGCGGAGGAGGAGGAGGAG 1769
 1491 AGAGGTGC----- 1498
 1770 TGAGGTCGCTGCTCGAGCACTGAATTTCTCAACAGTCTCTTGTCTGATGTTCTCT 1829
 1499 -----GGAAAGAG 1506
 1830 GTTTTGAACTTTTGTATGATACTAAGCTCGCTTAACTGTGGCCATTTGTTCAAGTTCAGGAAGAA 1889
 1507 CTTGATGACCTGAGAGACAGCTTCAGCTCTCAGCATGATGATGATGATGATGATGATGAT 1566
 1890 TTGAGCATCTGAGGAGCAAGCTGCGAGCGCTGACCATGATGATGATGATGATGATGATGAT 1949
 1567 AGAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1626
 1950 AGAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2009
 1627 GAGCGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1686
 2010 GAGCGGAGCGCGGATGAGCTGCGCGCTGTCGCGGAGCTCAAGTACGCGCGCTTCCAG 2069
 1687 GAAGTGAATCTGCAATTTGCCCACTTTGAAGGAGTCTTCTTGAAGAGATGATGATGATGAT 1746
 2070 GAAATCGAGCTCTATCTCCAGCTGGA---GAGCGAAACAGGGGAGAACCTGATGCTC 2126

1747 ACAGAAACGTTGGCCCTGAACACATTCCTG----- 1777
2127 ACCGAAACCGTCGGCCCTGAACAAATTCAGAGGTATGTTATTTCTTGTTCACCGTCA 2186
1778 ----- 1777
2187 CAAAAATTTTCAGAGCAAGTGCAGAAATTTGCCGATCGTCTCTAGTAGAGTAGTCTGTC 2246
1778 -----AGTTGTGAGCCGTTGGCAGGGATTCC 1805
2247 AGCGTGTGAATGTTGTTCTGTCCTATGCGCAGGTGTGAGCCGTTGGACGGGTATTC 2306
1806 AGTACAGAGACTTGGCCAAATAGAGAGGAGGTGATTGGTCTTGCTGATAGAGTTGCA 1865
2307 AGTGACCGGCTTGGCCAGACGCAAGAGAGAGGCTGGTTGGCCCTGGCTGACAGGCTTCA 2366
1866 TAAGCGGGTTGTGGACAGATCAAGCGGTAAATGCGAGTTTCTGAGGCAATTCAGAGTTC 1925
2367 CCAGAGGGTGTGGCCAGACAGAGAGGCTGTGAGCGCCCTGCGAGAGGCGGTGCTGAGGTC 2426
1926 AAGGACAGACTTGTAGGGCACACAGCCAACTGGATCATTTCTTATTCCTTGGACCAAC 1985
2427 GAGGCGCGGCTTGTGAGCGCCACACAGCCCACTGGCTCGTCTCTTCTTCTTGGGTCGAC 2486
1986 TGGTGTGGCAAACTGAGCTCGCCCAAGGCTCTTGTGAGCAGCTGTTTGTATGATGAAAA 2045
2487 TGGCGTGGGAAAACTGAGCTGGCCCAAGGCTTAGCCGAAACAGCTGTTTCGACGACGAGAA 2546
2046 CCTCTAGTTCGGAATGATGTGGAAATATAGGAACAACACTGTCTGTCTCGCCTCAT 2105
2547 CCTTCTTGTCCGCATCGCATGTGCGAGTACATGAGCAGCACTCGGTTGCGCCGCTCAT 2606
2106 TGGGCAACACAGCGGTATGTT----- 2127
2607 CGGAGCACCACTGGGTAAGTAGCAGAAAAATGCATGCTCTGCTATTTTAACTGGGAA 2666
2128 -----GGTCAAGAG 2136
2667 TTGCGATGAACCTGTTCTGACAGTGGCGCGACCTGTGACAGCTACGTCGGCCATGAA 2726
2137 GAAGGTGACAACTAACTAGGCTGTGAGGAGCGACCTTATTTGTGTCATCTTTTGTAT 2196
2727 GAGGTGGGCACTGACTGACAAAGTGTGAGGAGGCGGTACAGCGGTATCTCTTTCGAC 2786
2197 GAAGTGGAGAGGCTCATGTTGCTCTTCAACACTCTGCTCCAAAGTTTGGATGATGTT 2256
2787 GAGTTCGAGAGGCCCATGTGCGCGTGTTCACACCTTGCTCCAGGTCTCTCGACGAGGC 2846
2257 CGATTGACAGCGGCAAGGACGAGCACTCGAATTCAGGAACTCGGTGATTAATCATGACA 2316
2847 AGGCTGACGGATGGCAAGGACGAGCAGCGTGGACTTCAGGAAACCGGTGATCATCATGACA 2906
2317 TCARACCTTGGTGTGACACTCTTTCGAGGCTTAACCTGGGAAAGTAAACATGGAAGTG 2376
2907 TCGAACCTTCGGCGCGGACCTCTCTCGCTGGGATGGTGGGCAAGAACTCCATGAAGTTC 2966
2377 GCCCGGAGCTGTGTGATCGGG----- 2398
2967 GTCGCGATCTGGTCAATGAGGAGGTATGCATTCGATAACGTTCTGAATCGTGTGAGA 3026
2399 -----AGGTGAGGAA 2408
3027 CATTTGTTCTCTGACGCGTATGCTCAATCGGATCGTTATTTGCTGTGTGACAGGTGAGGAG 3086
2409 ACATTCACAGCAGAGCTTGAACAGGCTTACAGAGATTGTTGTTGTTGACCCCTTTC 2468
3087 GCATTTCCGCTTGGAGCTGCTGAACCGTCTGACAGATCGTGTATCTTCGATCTCTGTC 3146
2469 ACATGACAGGTTGAGAAAGTAGCTCGGCTTCAATGAAAGACGTTGCTGTCGGGCTTGC 2528
3147 CCACGAGCAGCTGAGGAGGTGCTCGCTTCAGATGAAGATGAGGATGAGGCGGCTGCTTGC 3206

QY 2529 TGAAGAGAGAGTGTCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTTGGCAGAG 2588
Db 3207 CGAAAGGGGCATCGCTCTGGCTGTGACCGAGCCCGCAATTGGACATCATCTTGTCTCTC 3266
QY 2589 TTATGACCC----- 2597
Db 3267 TTACGATCCGGTATGTGACCATCCATGATTTGATCCATCTGAATTCGTCGGTGACACCTG 3326
QY 2598 -----GGTGTATGTTGCTAGGC 2614
Db 3327 ATGGTGTGACTCTCTTATCTTTCTTGTGTGCTTCAACACAGGTGTATGCGCGCGGC 3386
QY 2615 CTATAAGGAGATGGATGGAGAAAGAGGTGGTGACAGAACTGTCAAGATGTTGTGCGTG 2674
Db 3387 CAATCAGGAGGTGGATTCGAGAAAGAGGTTGGTCAACAGCTGTGCGAAGATGCTGTATCCAGG 3446
QY 2675 AGGAATCGATCAAACTCCACTGTTTACATAGATG-----CAGGCGCTGGTGATCTTG 2728
Db 3447 AGGATGTCACAGAACTGTCACGGTCTCATCGCCGCGCCCGCAAGGACGAGCTGG 3506
QY 2729 TGTACCGGGTAGA---AAGTGGAGGTCTAGTGGACGCTTCAACAGGCAAGAGTCAAGTG 2785
Db 3507 TCTACAGGTGGACCGGAGCGCGGTCTGTTGAACGCTGAGACGCGGGATGAAGTCGGACA 3566
QY 2786 TGTGATTCATATGTTACCGGCCCAAGAGAGATGATGACAGCTCAGCGGTGGAAGA 2845
Db 3567 TCTGATCAGGTCCCAACAGCTCCACGAGGAGCGCTGCGCAGGCGCGTCAAGAAGA 3626
QY 2846 TGAGGATCAGAGAAATAGAGATGAGTAAATGAGGA 2882
Db 3627 TGAGGATCATGAGGAGGACGAGGACGCGCATGAGCA 3663

RESULT 4

US-09-070-927A-156
; Sequence 156, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Steven Barash
; Patrick J. Dillon
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 156:

SEQUENCE CHARACTERISTICS:

LENGTH: 6491 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 156:

99-070-927A-156

Very Match 21.7%; Score 672.8; DB 10; Length 6491;
 Local Similarity 59.1%; Pred. No. 3.4e-179;
 Matches 1180; Conservative 5; Mismatches 807; Indels 6; Gaps 2;
 683 AAGCAGGAGAGCTTGTATCTGTGATGCTGATGAGGAGATTAAGAGAGTCTGAGGGA 742
 546 AAGCAGGAGAGAGCTTGTATCTGTGATGCTGATGAGGAGATTAAGAGAGTCTGAGGGA 605
 743 TTCTTTCAGGAG 802
 606 TTTTATCAAG 665
 803 CAGCTGTGGTTGAAGGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 862
 666 CAGGATTTGTTGAAGGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
 863 CTGATGAGAGATTAATTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 922
 726 AAGATAAAG 785
 923 GAGAGTTTGAAG 982
 786 GCGAATTTGAAG 845
 983 TGATTTCTTTTATGATGAGATTCATTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1042
 846 TCATTTTATTTATGATGAGATTCATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 905
 1043 TGATGAGCTTATCTGTTCAAGCCAGTTAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102
 906 TGGACGCGGAGAGATTTATTAAG 965
 1103 CTACAGCTTTGAGAGATCAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162
 966 CCAACAGCTTTGAGAGATTTGAG 1025
 1163 TCCAAAGAGATTTAGTCAAG 1222
 1026 TCCAAAGAGATTTAGTCAAG 1085
 1223 TCAAGAGAGATTTAGGAG 1282
 1086 TAAAGAGAGAGATTTGAG 1145
 1283 CTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1342
 1146 CCCTACTTTATCCGATCGCTATTTACTGATCTTTTACAGAGAGAGAGAGAGAGAGAGAGAG 1205
 1243 TGGTTGATGAGGCTTTGCGGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1402
 1206 TAGTCGATGAAGCAAG 1265
 1403 ATAACTTTGAAG 1462
 1266 ATCAAGTAAACAGCTCGGTTAATGCMACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
 1463 AGGATAAGCAGCAGAGCTTATAG 1522
 1326 CAGATGATGAG 1385
 1523 ACAAGCTTACAGCTCTCAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
 1386 AAGAAGCAGCTCAATGAAG 1445

1583 GAAGGCTTAAACAG 1642
 1446 CCAATAAAGCTGCAGAGATTTGATAAAGCAAAACAGAGATTTAGAGATCTGAAAATAATT 1505
 1643 ATGACCTTGCAG 1702
 1506 ATGATTTAG 1565
 1703 TTGCCCAACTTGAAG 1762
 1566 TGAAGAAGATTGAG 1625
 1763 CTGAACAC ---ATTGCTGAGGTTGTGAGCCGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1819
 1626 CTGAAGACGAAATTTGCGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1685
 1820 GCCAAATTTGAG 1879
 1686 TTGAAGGCGAAACGAG 1745
 1880 GACAGAAATCAAGCGGTAAATGCAGTTTCTGAGGCAATTTAAAGGTCAGAGGCGAGAGAGAGAG 1939
 1746 GTCAAGATGAAGCCGTCGATGCTGTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
 1940 GTAGGCGCAACAG 1999
 1806 AAGATCCAAATCGCCCACTCGGTTGCTTCTTCTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1865
 2000 CTGAGCTGCCAAGAGCTTCTGCTGAG 2059
 1866 CAGAACTTGTCTAAAG 1925
 2060 TTGATATCTGCGGATATATGAAACAACTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2119
 1926 TTGACATGAGTGAATATGAG 1985
 2120 GGTATGTTGGTCAAG 2179
 1986 GCTATGTTGGTCAAG 2045
 2180 GTGCTATCTCTTTGATGAG 2239
 2046 CAATTTGCTTTATGAG 2105
 2240 AAGTTTTCGATGATGCTGATTTGAG 2299
 2106 AAGTTTTCGATGATGCTGATTTGAG 2165
 2300 CGGTGATATCATGATCAATCAAACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2359
 2166 CAGTTTAAATTTATGAG 2225
 2360 AAGTAAACAAT ---CGAGTGGCCCGGAGCTGCTGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2416
 2226 AAGGAG 2285
 2417 GACAGAGAGCTCTTGAACAGAGCTTGAAG 2476
 2286 AACAGAGAGCTTTTAAACAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2345
 2477 AGTTGAG 2536
 2346 ATGTGAAG 2405
 2537 GAGTTGCTTTGGCAGTCACTGATGCTGCTTGGAG 2596
 2406 AATTTGTTGAG 2465
 2597 CGGTGATGCTG 2656
 2466 CAGCTTATGAG 2525
 2657 CAAAGATGTTGTCGCTG 2674

2526 CAAAAGAAATTTGTTCTG 2543

|||||

10-398-221-5/c

Sequence 5, Application US/10398221

Publication No. US20040018514A1

GENERAL INFORMATION:

APPLICANT: KUNST, Frederik

APPLICANT: GLASER, Philippe

TITLE OF INVENTION: Listeria innocua, genome and applications

FILE REFERENCE: 344 702 - US

CURRENT APPLICATION NUMBER: US/10/398,221

CURRENT FILING DATE: 2003-03-27

PRIOR APPLICATION NUMBER: PCT/FR 01/03 061

PRIOR FILING DATE: 2001-10-04

PRIOR APPLICATION NUMBER: FR 00/12 697

PRIOR FILING DATE: 2000-10-04

NUMBER OF SEQ ID NOS: 4025

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 43980

TYPE: DNA

ORGANISM: Listeria innocua

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)..(end)

OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u

10-398-221-5

Query Match 21.5%; Score 668.2; DB 12; Length 43980;

Best Local Similarity 57.8%; Pred. No. 2.5e-177;

Matches 1239; Conservative 0; Mismatches 886; Indels 18; Gaps 2;

601 AAGAAGGGAAGAAAGTTGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTAAAGACT 660

10691 ACAGAGGGAAGAAAGTACTTCTCAAAATGACAGAGGAAGAACTATGAAGCTTTACAAA 10632

661 TATGAGAGATTTGGTTAGCAAGCA-----GGAGAGCTTGATTCCTGTGATTTGGTGT 714

10631 TATGACGAGATTTAGTCGCGAAGTAAGAAGCGGAAACTTGATCCAGTGAATGGACGC 10572

715 GATGAGGAGATTAGAAAGTCTGAGGAGTCTTTCAGGAGAACGAAGAACATCTCTGTG 774

10571 GACGAGAAATTCGTATGTATCCGAATTTTATCAAGAAACAAAAATTAATCCAGTT 10512

775 CTATGAGAGCCAGGAGTTGGTAAACAGCTGTGTGAGGTTTAGCAACAAGGATT 834

10511 CTAAATTTGGTGAACAGGTTGGTAAACAGCAATTTGGAGGTTTAGCAACAGTATT 10452

835 GTGAAGGAGATGTGCCACAGCTCTTACTGATGTGAGATTAAATTTGTTGGACATGGGT 894

10451 GTCAGAAAGACGTTCCAGAGAGGATTTGAAGATATAAACAATTTTCCCTTGATTTGGT 10392

895 CGCTTAGTTGCTGTGTCTAAATACCGAGGAGATTGAGAAAGGTTCAATCTGTTTTG 954

10391 TCCCTTAATGCTGGGCTAGTATCGTGGTGAATTTGAAGAACGTTTAAAGCAGTACT 10332

955 AAGAAGTTGAGGACGCTGAAGGCAAAAGTGAATCTCTTTATTTGATGAGATTCAATTTGGTT 1014

10331 CAAGAAGTAAACAAAGCGATGGACAAATTTGCTCTTTATAGATGAATTTACAAAT 10272

1015 CTGTGTGCTGCAAAACTGAAGGTCGATGATGATGATGATGATGATGATGATGATGATGAT 1074

10271 GTGCGCGAGGTAACAGAGTGGAGCTATGATGATGATGATGATGATGATGATGATGATGAT 10212

1075 GCTAGAGGCGAGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134

10211 GCAAGAGTGAACTTCACTGTATCGCGCAACAACTTTGACGAATACCCCAATATATC 10152

1135 GAGAAGATGCTGCTTTGAGAGGAGGTTCCAAACAAGTCTATGTTGCGGAGCCAAAGTGTG 1194

Db 10151 GAAAAGAGATGCTGCATAGAAAGACGTTTCCAAAAGTACTTGTTCAGAACCAACCGTG 10092

Qy 1195 CCTGACACCATTTAGTATCCTTTAGAGGACTCAGAGAGATGATGAGGACATCATGTGTG 1254

Db 10091 GAAGATACAGTTTCCATTTTACGTTGTTTAAAGAACGTTTGAATCCATCATGCGGTA 10032

Qy 1255 CGAATCCAAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTCTGTACATAAATCGGT 1314

Db 10031 AATATTCATGATATGCTTAGTTGCGCAGCTAGCTTCTTAATCGTTACATTACGAT 9972

Qy 1315 CGGATTTACCGGATAAGCAATTTGATTTGGTTGATGAGGCTTTGCGAAATGTGAGATC 1374

Db 9971 CGTTTATTTACCGGATAAGCAATTTGATTTAGTAGATGAAGCGTGTGCAACTATTTCGTGTC 9912

Qy 1375 CAGCTTCATAGTCAACCTGAAGAGATTGATACTTGAAGGAGGATTAAGCCAGCAAGCTCGACTTATAGAG 1434

Db 9911 GAAATTCATCTATGCGCAAGTGAACCTTGATGAAGTAAACAAGAAAGTCAATGACGCTAGAA 9852

Qy 1435 ATTGAATCTCACGCTTTGAAAGGAGGAGGATTAAGCCAGCAAGCTCGACTTATAGAG 1494

Db 9851 ATTGAAGAGCGCATTTAAAGAAAGAAAGAACCCAGCAAGTGAACGGCGCTTAGAGATG 9792

Qy 1495 GTGCGGAAAGAGCTTTGATGACCTGAGAGACAGCTTCAGCCTCTCAGATGAATACAGA 1554

Db 9791 TTACAGCGAAGCTAGCTGACTATTAAGAGAAAGCAATTAATAATGAAGTCTAAATGGAA 9732

Qy 1555 AAGGAGAAAGAGAGATTTGATGAGATTGCAAGGCTTAAACAGAAAAAGAGAGAGCTCATG 1614

Db 9731 TCCGAAAAAGCGAATTCAGTAAATTTGCGAAGTTCTGTAACAAATCGATCATCTGCGT 9672

Qy 1615 TTTTCTTTGAGGAGGAGCAAGCAAGATATGACTTTGCAAGAGCTGCTGATCTTAAGATAT 1674

Db 9671 CATGAATTTAGAGAGAGAGAGCAACTACGATTTAAATAAAGCTGCTGAACCTTGCCAC 9612

Qy 1675 GCGCAATTTCAAGAGTGGATCTGCAATTTGCCCACTTGAA-----GGAAT 1722

Db 9611 GGTAAATTTCCAGAGTAGAAAAAGAAATTTATTAGCAATTAGAAACCGAAAAATCCGAAAAA 9552

Qy 1723 TCTTCTGAAGAGATTTGATGCTTCAAGAAAAAGTTTGGGCTTGAACACATTTCTGAGGTT 1782

Db 9551 ACAGCCCAAGAGATCGAATTTTACAGAGAGAGTAAACAGAAATGAATTCGCTGAAT 9492

Qy 1783 GTGAGCGCTTTGAGACAGGATTCAGTGAAGAGCTTTGGCCAAAATGAGAAAGAGAGTTG 1842

Db 9491 GTTGGACGATGGAATGCAATGACAGTGAATAGTATAGAGAGAGAACCGCAAAACCTG 9432

Qy 1843 ATTGGCTTGTGATAGTTGCAATAGCGGTTGTTGGACAGAAATCAAGCGGTAATGCA 1902

Db 9431 CTAAATTTAGCCGATGTTCTTCAATAAAGTAAATTTGGTCAAGACGACGCGTTTCACTTA 9372

Qy 1903 GTTCTGAGGCAATTTCTAAGGTTCAAGGGCAGGACTTTGGTAGGGCACACAGCCAACTGGA 1962

Db 9371 GTTAGTATGATGATTTACGCTGCTGCGGGGTTTAAAGATCCAAACGCAATTCGA 9312

Qy 1963 TCATTTCTTATTTCTTTGGACCAACTGTTGTTGGCAAACTGAGCTGCGCAAGCTCTTGT 2022

Db 9311 TCTTTATTTCTTAGGCCCCAACCGGTTGTTGTTAAACCGAACTAGCCAAAGCATTAGCC 9252

Qy 2023 GAGCAGCTGTTTGTATGATGAATAAATCTTATTTCCGATTTGATGATGATGATGATGATGATGAT 2082

Db 9251 TATAATGTTTGTATTTCTGAGATCATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 9192

Qy 2083 CAACTCTGCTCTCGCTCATTTGGGGCACCAACAGGGTATTTGGTCAAGAGAGAGT 2142

Db 9191 AAACATTCGATCAAGACTTGTGCGGGCTCTCCAGGTTATTTGGATATGAGAGAGC 9132

Qy 2143 GGAACATTAATGAGGCTGTGAGGAGGAGCACTTATTTGTTGCTACTCTTTTGTATGAGATG 2202

Db 9131 GGAACATTTACGAAAGCTGTAAAGACCAACCAATTTTGAATTTGATTTGATTTGATTTGAT 9072

Qy 2203 GAGAAGGCTGATTTGCTGCTTCAACACTGCTTCCAGTTTGGATGATGATGATGATGATGATGAT 2262

Db 9071 GAAAAGCCATCCAGAGTATTTAATATCTTTTACAAAGTACTTGACGATGGTGGT 9012

2341407 CTTAAATAGCCGATGTTCTTCATCAAAAAGTAATGGTCAAGACGACCGGTTGAGTTA 2341348
 1903 GTTCTTCTGAGCAATTTCTAAGGTCAAGGCGAGCACTGGTAGGGCAACAGCACTGGA 1962
 2341347 GTTAGTGATGAGTATAGTGTCTGCGGGGATTAAGATCCAAACGACCAATCGGA 2341288
 1963 TCATTCTTATCTCTGACCAACTGCTGTGTGGCAAACTGAGCTGCGCAAGGCTCTTGCT 2022
 2341287 TCTTTTATTTCTTAGGCCCAACCGGTGTGTGTAACGGAACCTAGGCAAGCAATTAGC 2341228
 2023 GAGCAGCTGTTGATGATCAAAACCTTTAGTTCGGATGATATGTCGGAATATATGGA 2082
 2341227 TATATATGTTGATTTCTGAAGATCAATATGATTCGAATTTGATGTTGTAATACATGAG 2341168
 2083 CAACACTCTGTCTCTCGCTCATTTGGGGCACCACAGGAGTATGTTGGTCCAGGAAAGT 2142
 2341167 AAACATTTCCGATCAAGACTTGTGCGGGCTCTCCAGGTTATGTTGGATATGAGGAGC 2341108
 2143 GGACAACTAACTGAGCTGTGAGGAGCGACCTTATTTGTCATATCTCTTTGATGAAGTG 2202
 2341107 GGACAACTTTACCGAAGCTTAAGACGCAACCCATATTTGATTTGTTACTTGTGATGAATC 2341048
 2203 GAGAAGGCTCATGTTGCTCTCTTCAACACTCTGCTCCAGTTTTCGATGATGTCGATG 2262
 2341047 GAAAAGGCCATCCGACGATTTTAAATATCTTTTACAAAGTACTTTGACGATGCTCGGAT 2340988
 2263 ACAGCGGCAAGGCGAGGACAGCTCGATTTTCAGAACTCGGTGATATCATGACATCAAC 2322
 2340987 ACGGATTTCAAGGGCGCTTAATTCATTTTAAACACCGTAACTCATTTAGAGCTTAAT 2340928
 2323 CTTGGTGTGAACACTCTCTGAGGCTTCTGAGGCTTAACTGGGAAAGTAACTGGAAGTGGCCGG 2382
 2340927 ATCGGCTCTAAATTTTATTTACTTGAAGAACCGAAGAGGTGAAATCTCGCTGAACTAGAA 2340868
 2383 GACTGTGTGATCGGAGGTGAGGAAACACTTCAGACGAGCTCTTGAACAGGCTTGAC 2442
 2340867 TCAGAGGTAAGCAAAATTTTGAAGTGAATTTAAACAGAAATTTTAAACCGGGTAGAT 2340808
 2443 GAGATGTGTGTTTCCAGCCCTTTTCAATGACCACTGTTGAGGAAAGTACGCTCGGCTCAA 2502
 2340807 GATAATATTTCTATTTAAACCACTTACACTCGCTGATATTAAGGCAATTTGGAAGAAATTA 2340748
 2503 ATGAAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2562
 2340747 GTAGAGAACTTTCAAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340688
 2563 GCTTTGGACTATCTTGGCAGAGGTTATGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2622
 2340687 GCAAGAGCTTTTATTTGAGAGAGGCTTATGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340628
 2623 AGATGGATGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2682
 2340627 CGATATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340568
 2683 GATGAAACTTCCACTGTTTACATAGATGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2727
 2340567 ATGCCGCTATTTCTTCCGTTGAAATTTGATTTACAGATAAAGAAATTT 2340523

ULT 7
 10-369-493-43449
 sequence 43449, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 43449
 ; LENGTH: 2586
 ; TYPE: DNA
 ; ORGANISM: No. US20030233675A1loc punctiforme
 US-10-369-493-43449

Query Match 21.3%; Score 662; DB 12; Length 2586;
 Best Local Similarity 56.8%; Pred. No. 2.1e-176;
 Matches 1245; Conservative 0; Mismatches 935; Indels 12; Gaps 1;
 QY 531 CAGGATTTGTTGACCAAGTTCGTTAGCAGCGGAGGGTAAAGTTCGAGGTTGAGAA 590
 DB 372 CAAAGCTTTTATTCAGAGATTCGGTTTAGCAGAGGCAAACTAAAGATATATTAAACA 431
 QY 591 GCTTCGTGCGAAAGAGGAGAAAGTTGAGAGTCTTCAGGGGACACAAATTTTCAAGC 650
 DB 432 AGTTCGGGGAGCCAAAGAGTGACCGACCAAAATCCAGAGGCAAAATACGAGCACTGGA 491
 QY 651 TTTAAGACTTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTTGTCTGTGATGG 710
 DB 492 AAAATACGGCGTGAACCTCACAGAGCGCGCTAAAGGTCAACTCGATCCAGTGTGG 551
 QY 711 TCGTGTAGGAGATTTAGAGAGTCTGAGGATTTCTTCGAGGAGAACGAGAACAAATCC 770
 DB 552 GCGGATGATGAGATTCGCGCACTGTCCAAATTTCTGTCTCGCCGACCAAGAAATATCC 611
 QY 771 TGTCTTTATTTGAGAGCCAGGAGTTGTTAAACAGCTGTGTTGAAAGTTTAGCACAAG 830
 DB 612 TGTCTAATTTGTTGAACCGGCTTGTGTTAAACTGCTATCGTGAAGGATTAGCACAGG 671
 QY 831 GATTGTAAGAGAGATGTGCCCAAGCTTTACTGATGTGAGATTAATTTCTGTTGACAT 890
 DB 672 TATTAATGCAAGTGTATGATCCCAAGTCTTCAAGAGCCGCAAGCTAATTTCTTTAGATAT 731
 QY 891 GGTGCTGTTAGTTGCTGCTGCTTAAATACCGAGGAGGATTTGAGAGGTTTGAATCTGT 950
 DB 732 GGTGCTTTGATTCGCGGGGCAAAATTCGCGGGTGAATTTGAAGAACGGCTGAAAGCACT 791
 QY 951 TTTGAAGAGTTGAGAGCGCTGAGGCAAGGATTTCTCTTTTATTGATGAGATTCATTT 1010
 DB 792 ATTAAGAGTTACTGATCTGCGCGCAATATTTGTTTATTATTGATGAATTTCAAC 851
 QY 1011 GGTCTTGTGCTGCGCAAACTGAGAGGTCGATGAGTGCAGCTAATCTCTTCAAGCCCAT 1070
 DB 852 CGTTGTTGCGCTGTGTGCAACCCAGCGCGGATGATGCGGGTAACTTTGTAAGAACCGCAT 911
 QY 1071 GTTAGCTAGAGGCGAGCTTCGATGCAATTCGTCCTCAACGCTTGAAGATACAGGAATA 1130
 DB 912 GTTGGCGGGGTGATTCGCTGTATTTGGGCGGCACTCTAGATGAATACCGCAACA 971
 QY 1131 TGTGAGAAAGATGTGCTCTTTGAGAGGAGGTTTCAACAAAGCTTATGTTGCGAGCCCAAG 1190
 DB 972 TATCGAAAGGATGCGCACTAGAAAGACGCTTCCAGCAGGTTTATGTCGATCAACCTAG 1031
 QY 1191 TGTGCTGACACCATTTAGTATCTTTAGAGGACTCAAGGAGAGTATGAGGGACATCATGG 1250
 DB 1032 TGTAGAGATAGTATTTTCGATTTTTCGCGGGTTGAGAGAACGTTATGAAACCCACCGG 1091
 QY 1251 TGTGCGAATCCCAAGACAGAGCTTTTATAAATCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1310
 DB 1092 GGTAAATTTCTGATGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1151
 QY 1311 TGTGCGGCAATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGCGAATGTGAG 1370
 DB 1152 CGATCGCTTTTACCTGATTAAGCCATTGACTTGTGTAGCAAGCGCGCGCCAGATTAA 1211
 QY 1371 AGTCCAGCTTGTATGATCACTTGAAGAGATTTGATTAACCTTTGAAGAGAGAGATCGACT 1430

1212 AATGGAGATCACTCCAAACACAGAACTCGACGAATTTGATCGAGATTTGCAATT 1271
1431 GGAATTTGAACTTACCGCTTGGAAAGGAGGATGAAGCCAGCAAGCTCGACTTAT 1490
1272 GGAATGAGAGAGCTATCGCTGCAAAAAGAGCGATCGCGCTTCTCGTGAACCTTGA 1331
1491 AGAGTGGGAAAGAGCTTATGACCTGAGAGACAGCTTCAAGCTTCTCAGATGAATA 1550
1332 AAGACTAGAAAGAAATTTGCGATCTCAAGAGAGAAACAAAGAACCCCTTAATATCTCAATG 1391
1551 CAGAAGGAGAAAGAGAGAAATTTGATGAGATTCGAAGGCTTAAACAGAAAAGAGAGAGCT 1610
1392 GCAATCTGAAAAGATATCAATTTGACAAATTCATCCGTTAAAAGAGATGAACGGGT 1451
1611 CATGTTTCTTTGAGAGAGGAGAGAAAGAGATATGACCTTGCAGAGCTGCTGATTAAG 1670
1452 CAATTTAGAGATTCAGCAAGCAGAACCGATACGACCTTAAACCGAGCTCGGAGTTGAA 1511
1671 ATATGGGCGCAAT-----TCAGAGTGGAAATCTGCAATTCGCCCACTTGAAGG 1718
1512 ATACGGTAATTAAGTTGATGCTGATTCAGTTGGAAGCAGTAGAAGCTGAATTTGGCAAG 1571
1719 AACTTCTTCTGAAGAGAAATGTGATCTCACAGAAACGTTGGGCTCGAACACATTTGCTGA 1778
1572 TGCCAAAGAGTGGAAATCACTATACGGGAGAGAGTACAGAACTGATTTGCTGA 1631
1779 GGTGTGAGCGGTGGAGAGGATTCAGTACAGAGACTTGGGCCAAATGAGAGAGAG 1838
1632 AATTAATTTCTAAATGGACAGAAATTTCCCATCAGCAAGCTGTGGAATCTGAGAAAGAGAA 1691
1839 GTTGATTTGCTTCTCATAGTTGATAGTACGGGTTGTCGAGACAGATCAAGCGGTAA 1898
1692 ACTACTGCATTTAGAAATGAACTACGCCACCGTGTGATTTGGACAAAGAGACGCTAC 1751
1899 TGCAGTTTCTGAGCAATTTCTAAGGTCAAGGCGAGGACTTGGTAGGCGCAACAGCGCAAC 1958
1752 AGCGTAGCGATGCAATTCAGCGATCGCGCGTGGACTGGCGGATCCCAATCGTCCAT 1811
1959 TGGATCATTTCTTATCTTGGACCACTGTGTGTCGCAAACTGAGCTCGCCAGGCTCT 2018
1812 CGTAGCTTTATTTCTTGGGCTTACGGGTGTGGGTAAACCGAGTTGGCGAAAGCGCT 1871
2019 TGCTGAGCAGCTGTTGATGATGAACCTTTAGTTCCGATTTGATGATGTCGAATATAT 2078
1872 GCGGGCTATATGTTGATAGCAAGATGCGCTGGTGGCAATCGATGATGTCGAATATAT 1931
2079 GGAACAACTCTGTCTCTCGCTCATTTGGGCGACCAACAGGCTATGTTGGTCAAGCA 2138
1932 GGAGAAACACGCGCTCTCCGTTTAACTCGTGGCGCTCCAGGATATGTTGGTTACGAAG 1991
2139 AGTGGACAACTAACTGAGGCTGTGAGGAGCGACCTTATTTGTCATATCTTTGATGA 2198
1992 AGCGGACAACTAAACAGAGCGATTCGCGCGCTCTTACTCAGTGATTTCTTTTGAAGA 2051
2199 AGTGGAGAGGCTCATGTTGCTCTCTTCAACACTCTGCTCCAGTTTTCGATGATGCTG 2258
2052 AATCGAGAAAGCACCTTGAATTTTAAATCTTCTTGGAAATTTCTCGATGATGCTG 2111
2259 ATTGACAGCGGCAAGGAGGAGCAGTCTGATTTTCAGGAATCTGGTGATTAATCATGACATC 2318
2112 CGTCACTGATGCTCAAGGTCAUAAAGTGGACCTTCAGAAATGCTATTTATCATGACTAG 2171
2319 AAACCTTGGTGTGAACACTCTCTTCAGGCTAATCGGAAAGTAAACAAATGGAAGTGGC 2378
2172 CAACATCGGTTCCCAATATATTTCTTGTATGCTGGGAGTAATGCTCACTACGAGCAAT 2231
2379 CCGGAGCTGTGATCGGAGGTGAGGAAACACTTTCAGACAGAGCTTTGAAAGGCT 2438
2232 GCGCGCTGAGTCAAGAGCGATGCGGAATAGCTTTCGCTCAGAAATTTCTCAACCGGAT 2291
2439 TGAAGAGATTTGTTGTTCCAGCCCCCTTTTCAATGACCAAGTTGAGGAAAGTACGCTCGCT 2498
2292 TGACGAAATCATCATCTTCCACGGTTTAGATAAGAGGAATTCGCGCAGATTTGCTGTT 2351

QY 2499 TCAATCAAGACGTTGCTCGGCTTCTGCTGAAGAGAGGATTTGCTTTGGCAGTCACTGA 2558
Db 2352 GCAAGTACAAAGATTAGCCAAAGATTGCCCAATATCCCTCAAGCTCTCAGA 2411
QY 2559 TGTCTGTTGGACTATATCTTGGCAGAGAGTTATGACCCGGTGTATGCTGCTAGGCTAT 2618
Db 2412 TGTGCACTGACTTTTATAGCAGAGTATGATGATGATGATGATGATGATGATGATGAT 2471
QY 2619 AAGGAGATGGATGGAGAGAGGTTGTTGACAGACTGTCAAGAGATGTTGCTGAGGAG 2678
Db 2472 GAAACGGGCGATTCAGCGAGAGCTAGAACTCAAAATGCAAAAGCCATCTTGGCGGTGA 2531
QY 2679 AATCGATGAAACTCCACTGTTTACATAGATG 2710
Db 2532 ATTCACGATGGCAACACCATCTTTGTAGATG 2563

RESULT 8
US-10-369-493-42135
; Sequence 42135, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42135
; LENGTH: 2604
; TYPE: DNA
; ORGANISM: Lactococcus lactis
US-10-369-493-42135

Query Match 21.2%; Score 659.8; DB 12; Length 2604;
Best Local Similarity 58.1%; Pred. No. 8.9e-176;
Matches 1225; Conservative 0; Mismatches 872; Indels 12; Gaps 3;

QY 604 GAAGGAGAAAGTTGAGAGTCTTCAAGGGGACAAATTTTCAAGCTTTAAAGACTTAT 663
Db 439 GAGGTGATAAAGTGACCCAGTCAAAATGCAAGAGAAACATACAAAGCACTTGAATAAT 498
QY 664 GGAAGAGATTTGGTTGAGCAAG-----CAGGGAAGCTTGTATCTGCTGATTTGGTCTGAT 717
Db 499 GGGTAGATCTCGTTGCTCAAGTTAAATCAGTTAAACAGATCTGTCATTTGGACGTAT 559
QY 718 GAGGAGATTAGAGAGTGTGAGAGATTTCTTTCGAGGAGAACGAGAAACAATCTCTGCTT 777
Db 559 GAAGAAATTTCTGATGTCAATTCGAGTGTCTTCTCTGTAACAAACAAATAACCCGTTCTT 618
QY 778 ATTGAGAGCCAGGAGTTGTTAAACAGCTGTGTTGAGGATTTAGCAGAAAGATTGTG 837
Db 619 ATTGTTGAACCTCGGGTGTGTTAAACAGCCATTTGTTGAAGGATTTGGCAGAAAGATTGTC 678
QY 838 AAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTTCGTTGGACATGGGTGG 897
Db 679 AGAAAGACGTTCTGAAATCTGAAAGATAAACAATCTTTTCACTGTATATGGGTGCC 738
QY 898 TTAGTCTGCTGCTTAATACCGAGGAGAGTTTGAAGAAAGGTTTGAATCTGTTTGA 957
Db 739 CTGATTTCAGAGCGCAATATCTGTTGATTTTGAAGAGCTTTGAAGCAGTCTTAT 798
QY 958 GAGATTGAGACGCTGAGGCAAGGATGATCTCTTTATGATGAGATTCATTTGGTCTT 1017
Db 799 GAGTGAAATAATCAGATGAGCAAAATATCTCTTTTATGATGAACCTTCATACGATTTGT 858

QY	2095	TCTCGCCTCATTTGGGCGCACACAGGGTATTTGGTCTCAGAGGAAGGTGACAACTAACT	2154
Db	1939	TCTCGGTTAGTCGGAGGACCTCCAGATATATGTCGGTTACGATGAAGAGGCCAAATTGACT	1998
QY	2155	GAGGCTGTGAGGAGGGCGACCTTATTGTGTCATACTCTTTGATGAAGTGGAGAGGCTCAT	2214
Db	1999	GAAGCTGTTTCGTCGTAATCCTTATACAAATCATCTTGTCTTGATGAATCGAAAAGCACAT	2058
QY	2215	GTTCGTGCTTCAAACTCTGCTCCAAAGTTTGGATGATGGTTCGATTGACAGACGGGCAA	2274
Db	2059	CCAGATGTCTTCAATATCTTTGTCGAAGTTTGGATGACGGCGGTTTGACTGACTCTAAA	2118
QY	2275	GGCAGGACAGTCGATTTCTAGGAACTCGGTGATAATCATGACATCAAACCTTGGTCTGAA	2334
Db	2119	GGTGTCTCTCGTTGACTTTTAAGAACACAGTGTGTGATTGACTTCAAAATGTCGGTCTCAA	2178
QY	2335	CACCTCCTTTGACGGGCTAACTCGGAAAGTAAACAAT---GGAGTGGCCCGGCACTGTGTG	2391
Db	2179	TATTTGCTTTGATAATGTTGGAGAAATGTGAAATTCGGAAGAAACAACCTGAAACGTTG	2238
QY	2392	ATCGGGAGGTGAGGAAACACTTTCAGACACAGAGCTCTTGAACAGGCTTTGACGAGATTGTG	2451
Db	2239	ATGTCTCAACTTCGGGCAATTTCAAACCAGAAATTTCTAAATCGAATTGATACCACTT	2298
QY	2452	GTGTTCCGACCCCTTTTCATCAGCACCATGTTGAGGAAAGTAGCTCGGCTTCAAATGAAAGC	2511
Db	2299	CTTTTCAAACCATTGGCACTTGAGGATATTAAGATATTTATCGTTAAATGACTAGCCAA	2358
QY	2512	GTTCGTCTCGGCTTGCTGAAAGAGAGTGTGCTTTGGCAGTCACATGATGCTGCTTTGGAC	2571
Db	2359	CTTTTCACATCGTCTTTGAAGAAATGGAATGTTTCAACTTGAGTTGAGTGAAGAACTAAAAGTT	2418
QY	2572	TATATCTTGGCAGAGAGTTTATGACCCGGTGTATGCTGTAGGCTTATAGGAGATGGATG	2631
Db	2419	TGGATTGCGGAAAATGCTTTATGAACACAGCTTATGTTGCAACGTCCTCAAACGTTATTG	2478
QY	2632	GAGAGAGAGTGGTCACAGAACTGTCAAGATGTTGTTCGGTGAGGAAATCGATGAAAAAC	2691
Db	2479	ACAAAGTCATCGAAAATCCATTAGCTAGCTGATTTATCGGAGGAAAAAATCCACCAAAA	2538
QY	2692	TCCACTGTT	2700
Db	2539	TCAAAAGTT	2547

RESULT 9
US-10-310-154-117
; Sequence 117, Application US/10310154

Sequence 117, Application US/1031030030233670A1

Publication No. US20030233670A1

GENERAL INFORMATION:

APPLICANT: Edgerton, Michael D

APPLICANT: Chomet, Paul S.

APPLICANT: Adams, Thomas H

APPLICANT: Ruff, Thomas G.

APPLICANT: Agarwal, Ameeta K.

APPLICANT: Ahrens, Jeffrey E.

APPLICANT: Ball, James A.

APPLICANT: Ball, G.

APPLICANT: Bell, Erin

APPLICANT: Boddupalli, Raghava

APPLICANT: Deikman, Jill

APPLICANT: Deng, Molian

APPLICANT: Dong, Jinzhao

APPLICANT: Duff, Stephen M.

APPLICANT: Galligan, Meghan M.

APPLICANT: Hinchey, Brenda S.

APPLICANT: Huang, Shihshieh

APPLICANT: Johnson, G. Richard

APPLICANT: Jung, Vincent

APPLICANT: Kretzmer, Keith A.

APPLICANT: Laccetti, Lucille B.

APPLICANT: Lai, Chao-Qiang

APPLICANT: Lee, Gary

PPLICANT: Lin, Jie-yi
 PPLICANT: Liu, Jingdong
 PPLICANT: Lu, Bin
 PPLICANT: Luethy, Michael M.
 PPLICANT: Lund, Adrian
 PPLICANT: Madson, Linda L.
 PPLICANT: Malloy, Kathleen A.
 PPLICANT: McKiel, Christine L.
 PPLICANT: Miller, Philip W.
 PPLICANT: Padmavathi, Manchikanti
 PPLICANT: Parnell, Laurence D.
 PPLICANT: Start, William G.
 PPLICANT: Tennesen, Dan
 PPLICANT: Vidya, K.R.
 PPLICANT: Wang, Haiyun
 PPLICANT: Xin, Zhanguo
 PPLICANT: Xu, Nanfei
 PPLICANT: Yang, Chunzhi
 PPLICANT: Zeng, Xiaoping
 PPLICANT: Zhang, Qiang
 PPLICANT: Zhao, Yejuan
 PPLICANT: Zhou, Li
 TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
 FILE REFERENCE: 38-15(52796)B
 URRENT APPLICATION NUMBER: US/10/310,154
 URRENT FILING DATE: 2002-12-04
 RIOR APPLICATION NUMBER: 60/337,358
 RIOR FILING DATE: 2001-12-04
 UMBER OF SEQ ID NOS: 736
 Q ID NO 117
 LENGTH: 3000
 TYPE: DNA
 ORGANISM: Xylella fastidiosa
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (187)...(2769)
 OTHER INFORMATION:
 0-310-154-117
 396 TCCACCTGATGATATCCAGCAGTCTAGTCTTATTAAGGTCATTCGTGCTGCTCAAGC 455
 414 TCAGACGGTAAATGCTCACCAGCAATGAATTGAGTCGTTGTTTCATCGACTGATAA 473
 456 TGCTCAGAGTCAAGAGTGTGATCTCAATTTGGCTGTTGACCAAGTTGATTTATGGGCTTCT 515
 474 GTTAGCGCAGCAGATGGCGATCAGTTTATGCCAGTGAAGTGTGTTGGCTGTGGT 533
 516 TGAAGATTCT---CAAATCAGGATTTGTTGAACGAAAGTCGGTGTAGCGAGCGGAGGGT 572
 534 CGATGATAGTGGGGGGCTGGGCGAGCGTGGCTGCTGCTGGTGGGAAAAAAGAAAT 593
 573 AAAGTCTGAGTTGAGAGCTTCGTTGGGAAAGAGAGGAGAAAGTTGAGATGCTTCAAG 632
 594 TGAGGCTGGATGATATAATCGGTGGCGGTGAAACTGTCCAGACTGGAATGCGGAGGA 653
 633 GCACACAAATTTTCAAGCTTTAAAGACTTTATCGAAGAGATTTGGTTGAGCAAGCAGGAA 692
 654 GCAGCGTCAAGCGTTGGAAGATACAGATTGATCTGACTGCCAGGCTGAGATGGGAA 713
 693 GCTTGATCTGTGATGTTGGTGTGATGAGAGATTTAGAAGAGTCTGTGAGGATTTCTTCGAG 752
 714 GCTTGATCCGGTGAATGTTGGTGGGATGAGAAATTCGCGCAGATACAGGATTTGCAGCG 773
 753 GAGACGAGAGACAAATCTGCTCTTATTTGGAGAGCAGGAGTTGTTGAAACAGCTGTGGT 812
 774 GCGTCAAAAAATAATCTGCTGCTGATTTGGTGGAGCGGGTGTGGGTAATAAATGCGCATGT 833
 813 TGAAGGTTTAGCACAAGGATTTGTAAGAGGAGATGTGCCACAGCTTCTTACTGATGTGAG 872

Db 834 AGAAGGGCTGGCTCAGCGCATTTGTCAATGGTGTGAGGTTCCGGAAGGGTTGCGCAGTAAGCG 893
 Qy 873 ATTAATTTTCGTTGGACATGGGTGGCTTAGTTGGTGTCTAAATACCGAGGAGGATTTGA 932
 Db 894 TCTGCTCTCGCTGATTTGGAGAGGCTTGAATTTGGGGCCAAAGTTTCGTGGTGTGATTTGA 953
 Qy 933 AGAAGGTTGAAATCTGTTTGAAGAGGTTGAGGAGCTGAAGGCAAGGTTGATTTCTCTT 992
 Db 954 GAGCGCTTGAAGGGGGTCTTAACGATCTCGTAAGAAATGAGGGGGGGTCAATTTGTT 1013
 Qy 993 TATTGATGAGATTCAATTTGGTCTTGGTGTGGGAAACTGAAGGGTGTGATGTGAGC 1052
 Db 1014 CATTGACGAGCTGCATACCATGTTGGTGGCGGTAAAGCCGATGTGCGATGTGCTGG 1073
 Qy 1053 TAACTGTTCAAGCCCATGTTAGCTAGAGGCGACCTTCGATGCTATTCGTCTACCAAGCT 1112
 Db 1074 CAATATGCTCAAGCGGGTTAGCACTGTGTGATTTGATGTTATCGGTGGACTAGCTT 1133
 Qy 1113 TGAAGAAATACAGGAAATATGTTGAGAAAGATGCTTGCCTTTGAGAGGAGTTTCCAAAGT 1172
 Db 1134 GGATGAGTATCGCAAGTACATTTGAGAAAGGATCGCGCTTGAAGCGCGCTTCCAGAAAGT 1193
 Qy 1173 CTATGTTCCGAGCGCAAGTGTGCTGACCAATAGTATCTTTAGAGGACTCAAGAGAA 1232
 Db 1194 GTTTGTTGGGAGCGGACGCGTGGAGGATACCATCGCAATTTCTTCGTGAGCTTAAGGAAA 1253
 Qy 1233 GTATGAGGACATCATGTTGTCGAATCCCAAGACAGAGCTCTTATAAATCTGCTCAGCT 1292
 Db 1254 GTATGCGTTGCACCAAGGTTGGAATCACTGATCGGCTATTTGTTGCTCGGCTAGCT 1313
 Qy 1293 GTCTGCTGTTACATAACTGTTGGCTGGCAATTTACCGGATTAAGCAATTTGTTGTTGATGA 1352
 Db 1314 GTCTAATCGCTACATCATCTGATCGTACGTTTACCAGATAAAGCGATTGATTTGATGATGA 1373
 Qy 1353 GGCTTGTGCGAATGTGAGAGTCCAGCTTGTATGATGTCAACTGAAGAGATTGATTAACCTTGA 1412
 Db 1374 GGCTGCCAGCTGTTATTCGTATGGAATTTGACTCCAAGCCGGAAGAGCTTGTGTTGGA 1433
 Qy 1413 AAGAAAGAGGATGAGCTGGAATTTGAATTTCAAGCTTGGAAAGGAGAGGATTAAGC 1472
 Db 1434 GCGTGGCTGTTGTTGAGTTGAATTTCAAGCTGAGATGCTGAAGAAAGGATGAGGC 1493
 Qy 1473 CAGCAAGCTCGACTTATAGAGTGTGGAAAGAGCTTGTATGACCTGAGAGACAAGCTTCA 1532
 Db 1494 GAGCAAGCAGCTTTAGCGGATCTTGAAGCTGATATTGAGTTCTGCGACCGTGAATTTTC 1553
 Qy 1533 GCCTCTCAAGATGAATACAGAAAGAGAGAGAGATTTGATGAGATTGCAAGGCTTAA 1592
 Db 1554 CGATCTGAGAGAGTGTGGAGATCAGAAAGCCGCTACAGGGGGCGACTAAGATCAA 1613
 Qy 1593 ACAGAAAGAGAGAGCTCATGTTTCTTTTTCAGGAGGAGAGAGAGAGATATGACCTTGC 1652
 Db 1614 GGATCGATCGAGCAGCGCAAGCTTGAATTTGGAAGCGCGCAGCGCTCAGAGCTACGC 1673
 Qy 1653 AAGAGCTGTGATCTTAAGATATGCGCAATTTCAAGAGTGTGAATCTGCAATTTGCCAAT 1712
 Db 1674 CAAGATGAGCGAGATTGATGATGTTGTTCTCCCGGCTTGGAGAGCAGCTGTGTGGCAGC 1733
 Qy 1713 TGAAGGAATTTCTTGAAGAGAAATGATGATGCTCAGAAAGAGTTGGGCTTGAACAT 1772
 Db 1734 AAGTCAAGCGGAAACAGACAGATTTTACGTTGGTGGAGAGAAAGTACTGTGAGGAGAT 1793
 Qy 1773 TGCTGAGGTTGTGAGCGGCTTGGACAGGATTTCAAGTGAAGAGACTTTGGCCAAAATAGAA 1832
 Db 1794 TGCCGAGGTAGTCAAGCTGTGAGTCTGATTTCCAGTGAAGAGATGCTTGGGGCGAGCG 1853
 Qy 1833 GAGAGGTTGATTTGGTCTTGTGATAGTTTGAATTAAGCGGTTGTGGGACAGAACTCAAGC 1892
 Db 1854 CGATAAGTTTCTGCGTATGGAAGCTCATCTAGCGCCAGCTGTGTGTGGTCAAGAGAGGAGC 1913
 Qy 1893 GGTAAATGCAAGTTTCTGAGGCAATTTCAAGGTCAGGGCAGGACTTGTGTAGGGCACAA 1952
 Db 1914 GATCAAGGTGTGATCGATGCGGTACCGGCTTGGCTGCTGTTGTTGTTGATCTCAATCG 1973

JLT 10
110-369-493-41201
Sequence 41201, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEIN
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28

1293 GTCTGCTGTTACATACCTGCTGGCAATTTACCGGATAAAGCAATGATTGTTGATGA 1352
 1128 GTCTAATTCGCTACATCACTGATCGTCACTAGTTACCAGATAAAGCGATTGATGATGA 1187
 1353 GGCTTGTGCGAATGTGAGAGTCCAGCTTGTATGATCAACCTGGAAGATGATTAACCTTGA 1412
 11188 GGCTGCCAGTCTGATTCGTAATGGAATTTGATCTCAAGCCGGAAGAGCTTGTATGTTGA 1247
 1413 AAGGAAGAGGATGAGCTGGAATTTGAACCTTCACGCTTGGAAAGGAGAGATAAAGC 1472
 1248 GCCTCGCTGATTCAGTTGGAATTTAGCTGAGCTGAGATGCTGAAGAAAGAAAGATGAGC 1307
 1473 CAGCAAGCTCGATTTATAGAGTGGGAAAGAGCTTGTATGATCACTGAGAGACAAGCTTCA 1532
 1308 GAGCAAGCAGCTTTAGCCGATCTTGAGCGTGTATTTGAGGTTCTGGAACCGTGAATTTTC 1367
 1533 GCCTCTCACGATGAATATACAGAAGAGAGAGAGATAATTTGATGAGATTCGAGGCTTAA 1592
 1368 CGATCTGGAGAGGTGGAGATCAGAAAGCCGCTACAGGGGGGAGCTAAGATCAA 1427
 1593 ACAGAAAGAGAGAGCTCATGTTTCTTTGAGAGGAGAGAGAGAGAGATGACCTTGC 1652
 1428 GGAGTGCATGAGCAGCCAAAGCTTGTATTTGGAAGCCGCGCAGCCGCTCAGCACTACGC 1487
 1653 AAGAGCTGCTGATCTAAGATATGCGCAATTCAGAAAGTGGAAATCTGCAATTCGCCAACT 1712
 1488 CAAGATGAGCGAGATTCAGTATGCTGTACTCCCGGCTTGGAGAGCAGCTGTGGCAGC 1547
 1713 TGAAGAACTTCTTCTGAAGAGAAATGTGATGCTCACAGAAAGCTTGGGCTTGAACACAT 1772
 1548 AAGTCAGGCGGACAGACGATTTTACGTTGTTGTCAGAGAAAGTGAAGTCTGAGAGAT 1607
 1773 TGTGAGTTGTGAGCGTTTGGACAGGATTCAGTGACGAGACTTTGGCCAAATAGAGAA 1832
 1608 TGCCGAGGTAGTCAGTCTGAGTGGTGTATTCAGTGAGCAAGATCTTTGAGGGGAGCG 1667
 1833 GAGAGGTTGATTTGCTGCTGATAGTTGATAGCGGTTTGGGAGAGATCAAGATCAAGC 1892
 1668 CGATAAGTTTGGTATGAGAGCTGATCTAGCCGAGCTGTGGTTGGTCAAGAGAGAGC 1727
 1893 GGTAAATGCAATTTCTGAGCAATTTCTAGGCTCAAGGAGGAGCTTGTAGGCGCAACA 1952
 1728 GATCAAGTGTGATCGAGTGGGTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1787
 1953 GCAACTGAGATCACTTTATTCCTTGGACCAACTGTTGTTGGCAAACTGAGCTCGCCAA 2012
 1788 ACCGAGCGGCTCGTTCCTTTCTGGTCCGAGCGGTGTTGTAAGACTGAGTTATGTA 1847
 2013 GGTCTTGTGAGCAGCTGTTGATGATGAACCTTTAGTTCGATTTGATGATGTCGGA 2072
 1848 GGGCTGGCTGAATTTTGTGACAGTCAAGATGCAATGGTCCGCAATGATGATGATGA 1907
 2073 ATATATGAACAACACTCTGCTCTCGCTCAATTTGGGCGACACACAGGATGTTGGTCA 2132
 1908 GTTATGGAAGAGCACTTGTGGCGGCTGATTTGTTGGCTTCCGGCTATGTTGGTTA 1967
 2133 CGAGGAAGGTGAGCAACTTAACTGAGGCTGTGAGGAGCGACCTTTATTTGTCATACTCT 2192
 1968 TGAGGAAGGAGGTTATCTTACTGAAATGGTGGGAGCTCGGCTTACTCTCTGATCTTT 2027
 2193 TGATGAAGTGGAGAGGCTCATGTTGCTCTCTTCAACACTCTGCTTCAAGTTTGGATGA 2252
 2028 GGAATGAAGTGGAGAGGCGCATAGTATGTTTCAATATTTCTGCTGCAAGTACTTGA 2087
 2253 TGGTCAATGACAGAGCGGCAAGGAGGAGCTGCAATTTTCAAGAACTCGGTCATATCAT 2312
 2088 TGACGCTTGAAGTGGTCAAGGCGTACTGTTGATTTTCCATATACCGTTATTTGAT 2147
 2313 GACATCAAACTTGGTGTGAACACTCTTTCAGGGCTTAATGGGAAAGTAACTGGA 2372
 2148 GACATCGAACTGGGTTCACATCAGATCCAGAACTCAGTGGAGAGCGATTTCTCGGAG 2207
 2373 AGTGGGCC-----GGGACTGTGTGATGCGGAGGTGAGGAAACACTTCAGACACAGAGCT 2426

2208 GTATACGCAAAATGAAGCGCGCAGTGCATGAGGTGTGAGGGCAATTTTCGCCAGAGTT 2267
 2427 CTTGAAAGAGCTTGAAGAGATTTGCTGTTTCAAGCCCCCTTTCAATGACCAAGTTGAGAA 2486
 2268 CATTAAACAGGTTGAGATGATTTGTTGTTTTCACCCGTTGGAAGGCGCAATCAAGCA 2327
 2487 AGTACTCGGCTTCAAAATGAAGAGCTTGTCTCGGCTTGTGAAAGGAGGAGTTGCTTT 2546
 2328 GATTCGGGGATTTCACTGCGGGCTTGAAGAGCTCTCGCTGAATCTGAGTTGAAAT 2387
 2547 GGCAGTCACTGATGCTGCTTTGGACTATATCTTGCAGAGAGTTATGACCCGCTGTATGG 2606
 2388 AGATTTGATGATGATGCTGCTTGGAAATCTGTTGTAATCTCGGATTTGATCTGTATATGG 2447
 2607 TGCTAGGCTTATAGGAGATGATGGAAGAGAGGTTGTGACAGAACTGTCAAGAGAGTT 2666
 2448 TGCCCGCCGCTGGAAGAGAGGATCCAACTCTCAGCTGGAGATCCCTTAGCGCAGCAGAT 2507
 2667 TGTGCTGAGGAAATTCGATGAAACTCCACCTGTTTACATAGATCAGCGCTGCTGTATCT 2726
 2508 CTTAGCTGAGGCTTCTCAGTGGAGATCTGTGAGGTGGTGTGATGAGGAGGAGCT 2567
 2727 TGTGTAC 2733
 2568 TGTATTC 2574

RESULT 11
 US-10-369-493-32853
 ; Sequence 32853, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 32853
 ; LENGTH: 2571
 ; TYPE: DNA
 ; ORGANISM: Xylella fastidiosa
 US-10-369-493-32853

Query Match 21.0%; Score 650.6; DB 12; Length 2571;
 Best Local Similarity 55.5%; Pred. No. 3.6e-173;
 Matches 1300; Conservative 0; Mismatches 1034; Indels 9; Gaps 2;

396 TCCACCTGATGATATTCACGAGTCTTCTAGCTTATTATTAAGGTCACTGCTGCTCAAGC 455
 228 TCAGACGGTTAATGCTTCACCCAGCAATGAATGAGTCGTTTCTTCATCGGACTGATAA 287
 456 TGCTCAGAGTCAAGGATGATCACTATTGGCTGTTGACCACTGATTTATGGTCTTCT 515
 288 GTTAGCGCAGCAGCATGGTATCAGTTTATGTCAGTGGTTCGTTGCTGTGGT 347
 516 TGAAGATCT---CAATCAGGATTTGTTGAACAGAGTCGTTGTCGACCGCGAGGCT 572
 348 CGATGATAGTGGGGGCTGGGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 347
 573 AAAGTCTGAGTGTGAGAGCTCTGCTGGGAAGAGGAGAGAGATTTGAGAGTCTTCAG 632
 408 TGAGGCTGCGATTTGATTAATTTGCTGGCGGTGAACTGTCAGACTGAGATTCGGAGGA 467
 633 GGACACAAATTTTCAAGCTTTAAAGACTTATGAAGAGATTTGTTGTCAGCAAGCAGGAA 692

468 GCAGCGTCAAGCGTTGAAAGATACAGATTGATCTGACTGCCAGGCGTGAAGTGCAG 527
 593 GCTTGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 752
 528 GCTTGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 587
 753 GAGAACGAGAACCAATCTCTGCTTTATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 812
 588 GCCTACAAAATTAATCTCTGCTGATTTGATGAGCGCGGTGATGAGTAAATCTGCAATGT 647
 813 TGAAGGTTTACACAAAGAGATGTAAGAGAGATGTCGCCCAACAGTCTTACTGATGTGAG 872
 648 AGAAGGCGTGGCTCAGCGCATTTGTAATGATGATGATGATGATGATGATGATGATGATGAT 707
 873 ATTAATTTCTGTTGACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 932
 708 TCTGCTCTCGCTGATTTGGGGGGGTTGATTTGCTGGGGGCAAGTTTCTGGTGGATTCGA 767
 933 AGAAGGTTGAAATCTGTTTGAAGAGAGATGAGGAGCGCTGAAGGCAAGTGAATCTCTTT 992
 768 GGAGCGCTTGAAGGGGTGCTTAACGATCTCGCTAAGATGAGGGCGGTCAATTTGTT 827
 993 TATTGATGAGATTCATTTGTTCTTGGTGGTGGGCAAACTGAAGGGTGCATGATGATGAGC 1052
 828 CATTTGATGAGCTGCATACCATGTTGTTGGTGGCGGTAAAGCCGATGGTGGATGATGCTGG 887
 1053 TAATCTGTTCAAGCCCATGTTAGCTAGAGGGAGCTTGCATGATGATGATGATGATGATGAT 1112
 888 CAATATGCTCAAGCGCGGTAGCAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 947
 1113 TGAAGAAATACAGGAAATATGTTGAAGAGATGCTGCTTTGAGAGAGAGTTTCAAAAGAT 1172
 948 GGATGATATCGCAAGTACATTTGAGAGAGATGCGCGCTTGAAGCGCGTTTCCAGAGGT 1007
 1173 CTATGTTGCGAGCAAGTGGCTGACACCATTTAGTATGATGATGATGATGATGATGATGATG 1232
 1008 GTTTGTTGGGAGCGGAGGTTGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1067
 1233 GTATGAGGAGCATCATGTTGGTGGCAATCCAGACAGAGCTCTTATTAATGCTGCTAGCT 1292
 1068 GTATGCTGTCACACCGTGGTGGAAATCACTGATCCGCTATTTGTTGCTGGGCTAGCT 1127
 1293 GTCTGCTCGTTTACATACTGCTGGGCAATTTACCGGATAAGCAATTTGATTTGATGATG 1352
 1128 GTCTAATCGCTTACATCACTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 1187
 1353 GGCTTGGCGNATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1412
 1188 GGCTGCCAGTGGTATTCGTATGGAATTTGACTCCAAACCGGAAGAGCTTGTATGCTTTGGA 1247
 1413 AAGGAAGAGGATGAGCTGGAAATTTGAACTTTCAGCGCTTGGAAAGGAGAGATGAAGC 1472
 1248 GCGTGGCTGATTCAGTTGAAATTTAGCGGTGATGATGATGATGATGATGATGATGATGATG 1307
 1473 CAGCAAGCTGCACTTATAGAGTGGCGAAAGAGCTTGTATGATGATGATGATGATGATGATG 1532
 1308 GAGCAAGCAGCGTTTACTGCTATTTGAGCGGTGATTTGAGGTTCTGGACCGCTGAATTTTC 1367
 1533 GCCTCTCAGNATGAATACAGAAAGGAGAGAGAGATTTGATGATGATGATGATGATGATGAT 1592
 1368 CGATCTGGAGAGGTGGAGATCAGAAAAGCGCGCTACAGGGGCGGCTAAGATCA 1427
 1593 ACAGAAAAGAGAGAGCTCATGTTTCTTTGAGGAGGAGAGAGATGATGATGATGATGATGAT 1652
 1428 GGATGCTGATCAGCAGCGCAAGTTGATTTGGAAGCGCGGCGCTCAGAGCTACGC 1487
 1653 AAGAGCTGCTGATCTAAGATATGCGCAATTCAGAAAGTGAATCTGCAATTCGCCAACT 1712
 1488 CAAGATGAGCGAGATTCAGTATGTTGTTACTCCCGCGTTCGAGAAGCAGCTAGTGGCAGC 1547
 1713 TGAAGAACTTCTTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1772

1548 AAGTCAGCGGAAACAGCAGATTTTACTTTTGGTCAGGAGAAAGTGAATGCTGTGAGGAGAT 1607
 1773 TGCTGAGGTTTGAGCGCTTGGAAGAGGATTTCCAGTGAAGAGCTTTGGCCCAAAATGAGAA 1832
 1608 TGCCGAGGTAGTCACTGCTGGACTGGTATTTCCAGTGAAGAGATGCTTTGAGGGCGAGCG 1667
 1833 GGAGAGGTTGATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1892
 1668 CGATTAAGTTGTTGCGTATGGAAGCTGATCTAGGTGCGATGCTGCTGATGATGATGATGATG 1727
 1893 GGTAAATGAGCTTTCTGAGGCAATTTAAGGTCAAGGCGAGGACTTTGGTAGGGCAACA 1952
 1728 TATCAAGGTGATCGATGCGGTGCGGCTTTCGGTACTTGGTGTCTGATCCAAATCG 1787
 1953 GCCAACTGATCATTTTATTTCTTTGACCAACTGCTGCTTGGCAAACTGAGCTCGCCAA 2012
 1788 ACCGAGCGCTCGTTCTCTTTCTGGGTCGAGCGGCTGTTGGTAAGACTGAGTTATGATTA 1847
 2013 GGCTCTTCTGAGCAGCTGTTTGTATGATGAATAAACCTCTTAGTTGCGATTTGATGTCGGA 2072
 1848 GCGCTGCTGATTTTGTTCGACAGTCAAGATGCAATGGTCCGATTTGATGATGATGATG 1907
 2073 ATATATGAAACAACTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2132
 1908 GTTCAATGAAAAACATTTCTGCGCGGCTGATTTGCTGCACTCCGGGCTATGCTGGTTA 1967
 2133 CGAGGAAGGTGACAACTAACTGAGGCTGAGGAGGCGACTTATTTGCTGTCATCTCTT 2192
 1968 TGAGGAAGAGGTTATCTTACTGAATGCGCGAGCTCGGCTTACTCTCTGATTTCTTT 2027
 2193 TGATGAAGTGAAGAGGCTCATTTGCTGCTCTTCAACACTCTGCTCCTCAAGTTTGGATGA 2252
 2028 GGATGAAGTGAAGAGGCGCATAGTGTGTTCAATATTTCTGCTGCAAGTACTTGA 2087
 2253 TGCTGATGATGACAGCGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2312
 2088 TGGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2147
 2313 GACATCAAACTTGTGCTGCAACACTCTCTGCGGGCTAACTGGGAAAGTAAACAATGGA 2372
 2148 GACATCAAACTTAGGTTTCACTCAGATCCAAAGACTAGTGGAGAGGATTTCTCGGAGGT 2207
 2373 AGTGGCCC-----GGGACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2426
 2208 GTATACGCAATGAAGGCGGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 2267
 2427 CTGTAACAGGCTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2486
 2268 CATTAAACAGTTTGGATGATTTGTTTTCACCCGTTGGACCAAGGCGCAATCAAGCA 2327
 2487 AGTAGCTCGGCTTCAAAATGAAGAGCTTGTCTCGGCTTGTGAAAGAGGATTTGCTTTT 2546
 2328 GATTCGCGGATTTCACTCGGGGGCTAGAGAGGCTTCTGCTGATCTGATGATGATGATGAT 2387
 2547 GGCAGTCACTGATGCTTTTGGACTATATCTGAGGAGAGGATTTATGATCCCGGTGATG 2606
 2388 AGATTTGGATGATGCTGCGCTGGAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2447
 2607 TGCTAGGCTTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2666
 2448 TGCCCGCTCGCTGGAAGAGCAATCCAACTCAGCTGGAGATCCCTTTAGCGCGAGAT 2507
 2667 TGCTGCTGAGGAAATTCGATGAAACTCCCACTGTTTACATAGATGAGGCGCTGCTGATCT 2726
 2508 CTTAGCTGAGCGTTCTGCTGATGAGGATGATGATGATGATGATGATGATGATGATGATG 2567
 2727 TGT 2729
 2568 TGT 2570

quence 7037, Application US/09815242
tent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

Q ID NO 7037

LENGTH: 2571

TYPE: DNA

ORGANISM: Haemophilus influenzae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2571)

9-815-242-7037

ery Match 20.9%; Score 649.4; DB 9; Length 2571;
st Local Similarity 56.6%; Pred. No. 7.9e-173;
ches 1243; Conservative 0; Mismatches 946; Indels 6; Gaps 2;

426 TCTTATTAAGGTCATTCGTCTCAAGCTCTCAGAAAGTCACGAGGTGATCTCATTT 485

261 TATTAATCTACTTAATTTATGCGATAAATTCGACAGCAAAACCAAGATAAATTTATTC 320

486 GCGTGTGACCAAGTTGATTCGGTCTTCTTGAAGATTCCTCAAAACGAGGATTTCTTCAA 545

321 GAGCGAATTTGTTTTTTTGTTCAGCTTTAGAAAGACGAGAACGATCAGCGATATTTTCAA 380

546 CCAAGTCGGTGTAGCGACGCGAGGTAAGTCTCAGGTTGAGAGCTTCGTGGGAAAGA 605

381 AAGTTCGGTGTGAAAAGAAACAAATTTCCGACATTCAGACATTCAGAGGGGACA 440

606 AGGGAAGAAAGTTGAGAGTCTTCAGGGGACACAAATTTTCAAGCTTTTAAAGACTTATGG 665

441 AAACGTGAACGATCAAAATGCAAGAAAGACAGACAGCGCTTGAANAATATACATTTGA 500

666 AAGAGATTTGGTTGACAGAGGAGCTTGATCCTGTGATTTGTCGTGTGATGAGAGAT 725

501 TTTAAACCGCTCGTCAGAAAGTGGCAAACTTGATCCTGTAAATTTGGCGGTGATGAAGAAAT 560

726 TAGAAGAGTCGTGAGGATTTCTTCGAGGAGACGAGAACATCCTGTGCTTTATTTGAGA 785

561 TCGTCAGCCATTCAGATTTACACGTCGTACCAAAATTAACCTGTGTAAATTTGGTGA 620

786 GCCAGAGTTGGTAAACAGCTGTGTTGAAGTTTATAGCAAAAGGATTTGTGAAGAGAGA 845

621 ACCAGGTGTAGGAAAACGGCGATTTGTGAGGCTTTGCAAGCGCATTCGTAAACGGGCA 890

QY 846 TGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTGTTGACATGGGTGCTTAGTTGC 905
DB 681 AGTGCCAGAGGTTTGAANAATAAAGTGTCTTTCATTAATATGGGGGCTTGAATTC 740
QY 906 TGGTGTCTAAATACCGAGGAGAGTTTGAAGAAAGTTGAAATCTCTTTTGAAGAAAGTTGA 965
DB 741 TGGTGTGAAATATCGTGTGAATTTGAAGAAAGTTTGAAGAAAGTACTCAATGAATTC 800
QY 966 GAGCGCTGAAGGCAAGTGTCTTCTTATTTGATGAGATTCATTTGTTGTTCTTGTGCTGG 1025
DB 801 GAAAGAAAGGTGCGGTATCTCTTTTATTTGAGAAATTCATATCTATGTCGCGCGGG 860
QY 1026 CAAACTGAAGGTCGATGATGAGTCTCTGTTCAAGCCCATGTTAGCTAGAGGCA 1085
DB 861 TAAACCGATGTCGATGATGAGTCTCTGTTTGAAGAAAGTTTGAAGAAAGTTGCAAGGCG 920
QY 1086 GCTTCGATGATTTGGTCTTCAAGCTTTGAAGAAATATAGGAAATATGTTGAGAAAGTGC 1145
DB 921 ATTACATTTGCGTGGTGCATCTACTTTAGATGAATATCGTCAATATATCGAAAAAGATGC 980
QY 1146 TGCCTTTGAGAGGAGGTTTCCAAAGTCTATGTTCCGAGGCAAGTGTGCTGACACAT 1205
DB 981 CGACTTGAACCGCTTTTCCAAAGGTCCTTTTGAAGAAAGTGTGAGAAAGTATCCAT 1040
QY 1206 TAGTATCCTTAGAGGACTCAAGGAGAAAGTATGAGGACATCATGTTGTGCGAAATCCAAGA 1265
DB 1041 TCGGATCTTACGTGTTTGAAGAAAGCTTATGAATTTCAATCATCACTGATATTACTGA 1100
QY 1266 CAGAGCTCTTATAATGCTGCTCAGCTGTCTGCTGTTTACATTAATCTGCTGCGCATTTACC 1325
DB 1101 CCCAGCAATTTGCTGCTGAGCAAGCTTTTCAATCTGTTATATTTCCGATGCTCAGTTACC 1160
QY 1326 GGATAAAGCAATTTGATTTGTTGATGAGGCTTGTGCGAAATGTGAGAGTCCAGCTTGATG 1385
DB 1161 AGATAAAGCCATTTGATTTGATTCGATGAAGCAGCGCTTAGCATTCGTATGAAATAGATTC 1220
QY 1386 TCAACCTGAAGAGATTGATAACCTTTGAAGAAAGAGAGATGAGCTGCGAAATTTGAACCTCA 1445
DB 1221 TAAACCTGAACCGCTTCTGATCGTCTTGAACGCTGCTGTTATTTATCCATTTAAATTTGAACAACA 1280
QY 1446 GCGCTTGAAGGAGAGAGATTAAGCCAGCAAGCTGCTGATTTATAGAGTGGCGGAAAGA 1505
DB 1281 AGCGTTTCAAAAAGAGAGAGAGCAAGCAAGCTGCGAAACGTTTGAAGAAATGTTAGAGAAAGA 1340
QY 1506 GCTTGTAGCTGAGAGACAGCTTTCAGCTCTTCAAGATCAAGATCAAGAAAGAGAGAGA 1565
DB 1341 ATTGCTGAAAAGAACGTAATACGCGGAACTTGAAGAAAGTATGGAATCTGAAAGAGC 1400
QY 1566 GAGAAATGATGAGATTCGAAGGCTTAAACAGAAAGAGAGAGAGCTCATGTTTTCTTTTGA 1625
DB 1401 AACGCTTTCTGGCTCTCAACATATTAACAAAGAGTTAGATACCTGCAAAACCGAATAGA 1460
QY 1626 GGAGGAGAGAGAGATATGACCTTCCAGAGAGCTGCTGATCTAAGATATAGCGCAATTC 1685
DB 1461 ACAAGCTGCTGCGCGGGTGTATTTAGCGAAATTTGCTGAGTTGCAATATATGGCGCATCC 1520
QY 1686 AGAAGTGAATCTGCAATTTGCCCAACTTGAAGGAACTTCTTCTGAAGAGAAATGATGATGCT 1745
DB 1521 TGATCTTGAAGAAAGCACTTGAAGAGCTGAAACCGAGAGAGAAAGAAATGACGCTTTT 1580
QY 1746 CACAGAAACGTTGGGCTTGAACACATTTGCTGAGGTTGTGAGCGCTTGGACAGGAGTTC 1805
DB 1581 ACGCTATCGCGTCAAGATGAAGAAATTCGAGAAAGTCTTTCTTAAAGCCACAGGCAATTC 1640
QY 1806 AGTGAAGAGCTTGGGCAAAATGAGAGAGAGGTTGATGTTGCTTCTGCTGATAGGTTGA 1865
DB 1641 TGATCAAAATGATGAGAGAGGCGAGAGAGAAAGAACTTTTCTTAAAGCCACAGGCAATTC 1700
QY 1866 TAAGCGGTTTGGGACAGCAATCAAGCGGTAAATGAGTTCTGAGGCAATTTCTAAGGTC 1925
DB 1701 TAAACGAGTGTGCTCAAGAGAGAGCGGTTGATGCGGTAGCAAAACGCAATTCGCTGATG 1760
QY 1926 AAGGCGAGGACTTGGTAGGGCAACAAGCCAACTGATCATTTATTCTTCTTGGACCAAC 1985

```
LOCATION INFORMATION: (4747)..(4747)
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
```

VII. T 73

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c

FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Query Match 20.9%; Score 649.4; DB 12; Length 1830121;
Best Local Similarity 56.6%; Pred. No. 78-171;
Matches 1243; Conservative 0; Mismatches 946; Indels 6; Gaps 2;

QY 426 TCTATTAAAGTCAATTCGTCGTCCTCAAGCTCTCAGAGTCACGAGGTGATCACTCAITTT 485
DB 906639 TATTATCTACTTAATTTATGCGATAAATTCACAGCAAAACCAAGATAAATTTATTTTC 906698
QY 486 GCGTGTGACCAAGTTGATTTATGGTCTTCTTGAAGATTCCTCAATCAGGGATTTGTTGAA 545
DB 906699 GAGCGAAATTTGTTTTTTTGTTCAGACTTTTAGAAGAACGAGAACGATCAGCGATATTTGAA 906758
QY 546 CGAAGTCGCTGTAGCGCGCGGAGGTAAGTCTCAGGTGAGAGCTTCGTGGGAAAGA 605
DB 906759 AAAGTCGCTGCGAAAGAAAGAAACAAATTCGCAAGCTATTACGCACTTAGAGGGGACA 906818
QY 606 AGGGAAGAAAGTTGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGG 665
DB 906819 AAACGTGAACGATCAAAATGCAAGAAAGAAAGCAGCAAGCGCTTGAAAAATATACGATTGA 906878
QY 666 AAGAGATTTGGTTGAGCAGCAGGGAAGCTTCTGATTTGATTTGCTGATGAGGAGAT 725
DB 906879 TTTAACCGCTCGTGCAGAAAGTGGCAAACTTGTATCTGTAATTGGCGCTGATGAAGAAAT 906938
QY 726 TAGAAGAGTCGTGAGGATTTCTTCAGCAGAACGAGAACCAATCTCTGTCTTATTGGAGA 785
DB 906939 TCGTGAAGCCATTCAAGTATTACACGCTGATCCAAAATTAACCTGTGTTAATTTGGTGA 906998
QY 786 GCCAGAGTTGGTAAAAACAGCTGTGGTTGAAGGTTTAGCACAAGAGGATTTGTAAGGAGA 845
DB 906999 ACCAGGTGTAGGAAAAACCGGATTTGCGAAGGCTTGCCACAGCGCATCGTAAACGGCA 907058
QY 846 TGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTGTTGACATGGGTGCTGTAGTTGC 905
DB 907059 AGTGCCAGAAGGTTTGAATAAATAAAACGTTCTTCAATTAGATATGGGGCGCTTGAITTC 907118
QY 906 TGGTGTCTAAATACCGAGGAGAGTTTGAAGAAAGGTTGAAATCTGTTTTTGAAGAGGTTGA 965
DB 907119 TGGTCCGAATATCGTGGTGAATTTGAAGAACGTTTAAACAGTACTCAATGAACTTTC 907178
QY 966 GGACGCTGAAGGCAAGTGTCTCTTTTATTGATGAGATTCATTTTGGTTCTTGGTGTGG 1025

908259	DB	TGCGATGGTGGCTATTGATATGTGAGAGTTTATGGAAAAACA	CAGTGTGTTCTCGTTTAGT	9083118
2106	QY	TGGGSCACCCAGGGTATGTTGGTACAGGAGGTGGACAACTAACT	TGAGGCTGTGAG	2165
908319	DB	TGGTGGGCTCCAGGCTATGTGGCTATGAAAGGGCGTTATTTAACT	TGAAGCTGTTGG	908378
2166	QY	GAGCGACCTTANTTGTGTCTATCTCTTTGATGAAGTGGAGAAAGCTCATGTT	TGCTGTCTT	2225
908379	DB	TGCTCGTCCATATTTCAGTGTATCTTATTAGATGAAGTTGAAAAGCACACG	CAGATGTATT	908438
2226	QY	CAACACTCTGCTCCAACTTTTGGATGATGTCATTCAGACAGCGGCAAGG	CAGACAGT	2285
908439	DB	CAATATCTTATTCAAGTGTGGATGATGTCGTTTAACTGATGTC	CAAGTCGACTGT	908498
2286	QY	CGATTTCAGGAAGCTCGGTGATATCATGATCAATCAAACTTGGTGTG	ACACCTCTCTGC	2345
908499	DB	GAATTCGTTAACTCTGTGTTATTATGACCTCTAACTTGGGTTCTGATTT	TAATCCAAAG-	908557
2346	QY	AGGGCTAACTTGGGAAAGTAAACAATGGAAGTGGCCCGGACTGTGTGAT	CGGGAGGTGAG	2405
908558	DB	-GTAATAAGACGAAAGCTATAGGAAA-----TGAAAGCCTTAGTGA	TGTCAGTGGTAAG	908612
2406	QY	GAACACTTCAGACGAGCTCTTGAAACAGGCTTGACGAGATGTGGTGT	TGCGACCCCT	2465
908613	DB	CCAACTTTCCGCCCAAGATTCATCAACCGTATTGACGAAACCGTGGTAT	TCCATCCACT	908672
2466	QY	TTCACTAGCACAGTTGAGGAAAGTACCTCGGCTTCAATGAAAGAGGTGCT	GTCTCGGCT	2525
908673	DB	TGGTAAAGAAATATCTCGTGGATTCGAAGTATCCAAATTAGAACCGCT	TAGCAAAACGTAT	908732
2526	QY	TGCTGAAAGAGGAGTTGCTTTTGGCAGTCACTGATGCTGCTTTTGGACTAT	TATCTTGGCAGA	2585
908733	DB	GGAAACTCGTGGATACGAATTGGTGTGTACCAATGCTTTATTAGATCTT	CATTATGGCGAAGT	908792
2586	QY	GAGTTATGACCCGGTGTATGTCGTAGGCTATAA	2620	
908793	DB	GGATACGACCAAAATTTATGTCGACGTCCATTTGA	908827	

RESOL 17
US-10-329-960-1

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c

```

```

FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```

```

NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Query Match          20.9%; Score 649.4; DB 15; Length 1830121;
est Local Similarity 56.6%; Pred. No. 7e-171;
atches 1243; Conservative 0; Mismatches 946; Indels 6; Gaps 2;

426 TCTTATTAAAGTTCATTCGTCTGCTCAAGCTGCTCAGAGTCACGAGTGTACTCATTT 485
906639 TATTAACTCTAATTATTTATCGATAAATTCGCACAGCAAAACCAAGATAAATTTATTTC 906698

486 GGCTGTTGACCAGTTGATTATGGTCTTCTTTGAAGATTCTCAAAATCAGGAGTTTGTGAA 545
906699 GAGCGAATTGTTTGTTCAGCTTTAGAAGACGAGGAACGATCAGCGATATTTTGA 906758

546 CGAGTCGGTGTACGACGGCGAGGTTAAAGTCGAGTTTGAAGAGTTCGTGGGAAAGA 605
906759 AAAGTCGGTGTGCAAAAAGAACAAATTTTCGACAGCTATTTCAGCAATTAGAGGGGACA 906818

606 AGGGAAGAAAGTTCAGAGTGTCTTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATCG 665
906819 AAACGTGACGATCAAAATCGAGNAGAAAGACGACAGCGCTTGAAAAATATACGATGA 906878

666 AAGAGATTTGGTTGAGCAAGCAGGGAAGCTTGATCCTGTGATTTGGTTCGTATGAGGAGAT 725
906879 TTTAACCGCTCGTCAGAAAGTGGCAACTTGATCCTGTAAATTTGGCGCTGATGAAGAAAT 906938

726 TAGAAGAGTCTGAGGATTCCTTCGAGAGAACGAAAGCAACATCCTGCTTATTGGAGA 785
906939 TCCTCGAGCCATTCACGATTTACACGTCGTACAAATAATCACCTGTGTTAATTGGTGA 906998

786 GCCAGGAGTTGGTAAAAACAGCTGTGGTTGAAGGTTTAGCAACAAGAGTTGTGAAAGGAGA 845
906999 ACCAGGTGTAGGAAAAACGGCGATTTGTCGAAGGCTTGGCACAGCCGATCGTAAACGGCA 907058

846 TGTCCCAACAGTCTTACTCATGTGAGATTAAATTTCTGTGACATCGGTGCGTTAGTTGC 905
907059 AGTCCAGAGAGTTTGAAAAATAAACGTGTGCTTTCATTAGATATGGGGCGCTTGATTGC 907118

906 TGGTGCTAAAATACCGAGGAGATTGAAAGAAAGGTTGAAATCTGTTTTCGAAAGAGTTGA 965
907119 TGGTGCGAATATCTGTGTGAATTTGAAGAACGTTTAAAGCAGTACTCAATGAACCTTC 907178

966 GGAAGCTGAAGGCAAGTGAATCTCTTTATTTGATGAGATTCATTTGGTTCCTGGTCTGG 1025
907179 GAAAGAGAAGAGTTCGCTTATCTCTTTATTTAGCAAAATTCATCTACTGTCGGCGCGGG 907238

1026 CAAAACTGAAGGGTTCGATGATGACGCTAAATCTGTTTCAAGCCCATCTGTAGCTAGAGGCA 1085
907239 TAAAACCGATGTCGATGATGCGGTAAATTTGTTTAAACCAAGTTTGGCACAGGCGCA 907298

1086 GCTTCGATGCAATCGTGCTACACGCTTGAGAAATACAGAAATATGTTTGAAGAAATGTC 1145

```

907299	ATTTACATTCCGGTGGCACTACTTTTAGATGAATATATCGTCAATATATCAAAAAAGATGC	907358
1146	TGCTTTTGAGAGAGGTTCCAAACAGTCTATGTTGGGAGCGAAGTGTGCTGCACCAAT	1205
907359	CGCACTTGAACGCCGTTTCCAAAAAGTCTTTGTGGACGAACCAAGTGTAGAATACCAT	907418
1206	TAGTATCCTTTAGAGACTCAAGGAGAGTATGAGGGACATCTGGTGTGGAAATCCAGA	1265
907419	TGCGATCTTACGTGGTTTGAAGAAGCTTTATGAAATTCATCATCACGTGGATATTACTGA	907478
1266	CAGAGCTCTTATAAATGCTGCTCAGCTGCTGCTGCTTTACATAACTCGTCCGGATTATCC	1325
907479	CCACAGCAATGTGCGTGCAGCAACGCTTTACATCGTTATATTTCCGATCGTCAGTTACC	907538
1326	GGATAAGCAATTTGATTTGGTTGATGAGGCTTTGTGCGAATGTGAGAGTCCAGCTTGATAG	1385
907539	AGATAAAGCCATTGATTTTATCGATGAAGCAGCGCTTAGCATTTGATTTGGAATTAGATTTC	907598
1386	TCAACCTGAAGAGATTGATAACCTTTGAAGGAAGAGATGCAGCTGGAAATTCGAATTCGA	1445
907599	TAAACCTGAACCGTTTGATCGTCTTTGAACGTCGTATTATCCATTTAAATTTGGAAACAAC	907658
1446	CGCCTTTGAAAAGGGAAGGATAAAGCCAGCAAAAGCTCGACTTTATAGAGTGGCGAAAG	1505
907659	AGCGTTTACAAAAGAAGAAGACGAAGCAAGTCGCAACGTTTGAATAATGTTAGAGAAAGA	907718
1506	GCTTGATGACCTGAGAGACAAGCTTCAGCCTCTCACGATGAATACAGAAAGAGAGAAGA	1565
907719	ATTGGCTGAAAAGAACGTGTAATCGCCGAACCTTGAAGAAAGTATGAAATCTCAAAAAGC	907778
1566	GAGAAITGATGAGATTCGAAGGCTTAAACAGAAAAGAGAAGAGCTCATGTGTTCTTTTGCA	1625
907779	AACGCTTTCTGCTCTCAACATATTAACACAGAGTTAGATACTGCAAAAACCGAACTGAG	907838
1626	GGAGGCAGAAACGAAGATATGACCTTGCAGAGCTGCTGATCTTAGATATGCGCCCAATTCGA	1685
907839	ACAAGCTCGTCCGCGGGTGATTTAGCGAAATGCTCTGAGTTGCAATATATGCGCGCATCC	907898
1686	AGAAGTGGGAATCTGCAATTTGCCCACTTTGAAGGAACCTTCTCTGAAGAGAAATGTGATGCT	1745
907899	TGATCTTTGAAAAGCACTTTGAGCAAGCTGTAACACCGAGGAAAGAAATGACGCTTTT	907958
1746	CACAGAAAAAGTTGGGCTCGAACACATTTGCTGAGGTTGTGAGCGCTTGGACAGGGATCC	1805
907959	ACGCTATCGCGTCAAGATGAAGAAATCGCAGAAAGTCTTTCTAAAGCCACAGGCATTCG	908018
1806	AGTGACGAGACTTTGCCAAATATGAGAGGAGAGGTTGATTTGTTCTGCTGATAGTTGCA	1865
908019	TGTATCAAAAATGATGGAAGCGGAGAAAGAAAACCTCTTGCATATGGAAGATCACTACA	908078
1866	TAAAGCGGTTTGGGACAGAATCAAGCGGTAAATTCGCAAGTTTCTTGAGCAATTTCTAAGGTC	1925
908079	TAAAACGAGTGATTGTTCAAGAAGAAGCGGTTGATGCGGTAGCAAAAACGCGATTCTGCTAG	908138
1926	AGGGCAGCACTTTGTAGGGCACAAACGCCAACTGGATCACTTCTTATCTCTTGGACCAAC	1985
908139	TGTTGACGCTCTTTCCGATCTTAATTCGCCCAATTTGGTTCTTTCTTCTCTAGGGCCAAC	908198
1986	TGTTGTTGGCAAACTGAGCTTCGCCAAGGCTCTTCTGCTGAGCAGCTGTTTGTATGATGAAAA	2045
908199	AGGTGTTGGGAAAAACAGAGCTTTGCAAAACTTTTGGCTAAATTTCTTTGTATGATGAAGA	908258
2046	CCTCTTAGTTTGGATTTGATATGTGCGAATATATGGAACAACAATCTGTCTCTCGCCTCAT	2105
908259	TGCCATTTGGTTCGATTTGATATGTTCAGAGTTTATGGAAAAAACACAGTGTCTCTCGTTTGT	908318
2106	TGGGGCACCAACAGGTTATGTTGTTCACGAGGAAGTGGACAACCTAACTGAGGCTGTGAG	2165
908319	TGGTGGCCTCCAGGCTATGTGCGGTATGAGAAAGCGGTTATTTAACTGAGAGCTGTTCCG	908378
2166	GAGGCGAAGCTTATTGTGTCACTCTTTTGTATGATGAGTGAAGGCTCATGTTGTGCTCTT	2225
908379	TGGTGGTCCATATTCAGTGACTTTATTAGATGAAGTTGAAAAAGACACAGCATGATGATT	908438

2226 CAACACTCTCTCCAGTTTGGATGATGTCGATTCAGACGGGCAAGCAGGACAGT 2285
 908439 CAATATCTTATTACAGGTGTTGGATGATGTCGTTTAACTGATGATGTCGACTGT 908498
 2286 CGATTTCAGGACTCGGTGATATCATCAACCTTGGTGTGAGACACCTCTCTTGC 2345
 908499 GGAATTCGCTAAGCACTGTGTTATTATGACCTCAACTTGGGTTCTGATTAACTCAAG- 908557
 2346 AGGCTAACTGGGAAAGTAACAATGGAAGTGGCGGAGCTGTGATGCGGAGGTGAG 2405
 908558 -GTAATTAAGACGAACTATAGGAAA-----TGAAAGCCTTATGATGTCAGTGTGTAAG 908612
 2406 GAAACACTTCAGACCGAGCTCTTGAACAGGCTTGACGAGTATGTCGTTTCCACCCCT 2465
 908613 CCAACATTTCCGCGCAAGATTCATCAACCGTATTGACGAAACCGTGTGTTATCCATCCACT 908672
 2466 TTCACATGACCGTGTGAGAAAGTATGTCGCTGCTTCAATGAAGAGCTGTGTCGTCGGCT 2525
 908673 TGGTAAGAAATATCCGTCGATTCGAGTATCAAGTATCAATTAAGAACGCTTACGAAACGAT 908732
 2526 TGCTGAAGAGGAGTGTCTTTGGCAGTCACTGATGCTTTCGACTATATCTTGGCAGA 2585
 908733 GGAACTCGTGGATACGAATGCTGTTTACCGATGCTTTATTAGACTTCAATGGCGAAGT 908792
 2586 GAGTTATGACCCGGTGTAGGTCAGTCTAGGCTATATA 2620
 908793 GGGATACGACCCCAATTTATGTCGACGCTCCATTGA 908827

LT 15
 0-369-493-33061
 Sequence 33061, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 33061
 LENGTH: 2571
 TYPE: DNA
 ORGANISM: Xylella fastidiosa
 0-369-493-33061

ery Match 20.9%; Score 649; DB 12; Length 2571;
 est Local Similarity 55.4%; Pred. No. 1e-172;
 tches 1299; Conservative 0; Mismatches 1035; Indels 9; Gaps 2;
 396 TCCACCTGATGATATCCAGCGAGTCTTAGTCTTATTAAGTTCATTCGTCGTCACAGC 455
 228 TCAGACGGTTATGTCCTACCCAGCAATGAATGAGTCGTTGTTTTCATCGGACTGATTA 287
 456 TGCTCAGAGTCAACGAGTGATCTCAATTTGGCTGTGTGACCAAGTTCGATTTATGGGCTTCT 515
 288 GTTAGCGCAGCAGCATGGCGATCAGTTTATGGCCAGTGAATGATGTTGCTGTGCTGTGT 347
 516 TGAGATTTCT--CAATCAGGATTTGTTGAACGAGTTCGGTGTAGCAGCGCGAGGCT 572
 348 CGATGATAGTGGGGGGCTGGCGCAGCGTGTGCTGTGCTGTGCGGAAAGAAAGAAAGAT 407
 573 AAAGTCTGAGTTGAGAGCTTCGTCGGAAAGAGGAGGAGAAAGTTCGAGTGTCTCAGG 632
 408 TGAGGCTGCGATTGATTAATTTGCTGGCGGTGAAACTGTCCAGACTGAGAAATGCTGAGGA 467

QY 633 GGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAA 692
 DB 468 GCGCGCTCAGCGTTGGAAAAGTACACGATTGATCTGACTGCGAGGCTGAGAGTGGCAA 527
 QY 693 GCTTGATCTGCTGATGTCGTCGATGAGGAGATTAGAAGAGTTCGTGAGGATTTCTTTCCAG 752
 DB 528 GCTTGATCCGCTGATGTCGTCGATGAGGAAATTCGCCGACGATACAGGTATTGCGACG 587
 QY 753 GAGAACGAGAAACAATCTCTGCTTTATTTGAGAGCCAGGATTTGGTAAAAACAGCTGTGGT 812
 DB 588 CGGTACAAAAAATACTCTGCTGCTGATTTGTCGAGCCGGCTGTGGTAAAAATGCCATTGT 647
 QY 813 TGAAGCTTTAGCACAAAGATTTGAAAGGAGATGTGCCAACAGTCTTACTGATGTGAG 872
 DB 648 AGAAGCTTGGCTCAGCGCATTTGCAATGGTGAGGTTCCGAAAGGTTGCGCAGTAAGCG 707
 QY 873 ATTAATTTCTGTTGACATGCGGTGCTTAGTTGCTGCTTAAATACCCAGGAGAGATTGA 932
 DB 708 TCTGCTCTCGCTGATTTGGGGCGCTTGATTTGCTGGGSCAAGTTTCTGCTGAGTTTGA 767
 QY 933 AGAAAGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGAAGGCAAGATGATTTCTTT 992
 DB 768 GGAGCGCTTGAAGGGGCTGCTTAAACGATCTCGCTAAGAAATGAGGGCGGGCTCAATTTGTT 827
 QY 993 TATTGATGAGATTCATTTGGTTCTTTGGTCTGGCAAAACTCAAGGGTCTGATGGATGCGC 1052
 DB 828 CATTGACCGCTGCATCACTGTTGGTGGCCGTTAAGCCGATGGTGGGATGGATGCTGG 887
 QY 1053 TAATCTGTTCAAGCCCATGTTAGCTAGAGGSCAGCTTCGATGTCATTTGGTGTCTACAAGCT 1112
 DB 888 CAATATGCTCAAGCGCGCTTAGCACGCTGGTGAGTTGCTTGTATCGTGGCTGCTACGTT 947
 QY 1113 TGAAGAAATACAGGAATATGTTGAGAAAGATGCTGCCTTTGAGAGGAGTTTCCAAAGT 1172
 DB 948 GGATGAGTATCGCAAGTACATTGAGAAAGATGCGCGCTTGGAAACGCCCTTTCCAGAAAGT 1007
 QY 1173 CTATGTTGCGGAGCAAGTGTGCTGACACCATTAAGTATCTTAGAGACTCAAGAGGAA 1232
 DB 1008 GTTTGTTGGGAGCGAGCTGGAGGATCCATCGCAATTTCTGTCGACTTAAGGAAA 1067
 QY 1233 GTATGAGGGAATCATGCTGTGCGAATCCAAAGACAGAGCTCTTTATAAATGCTGCTCAGCT 1292
 DB 1068 GTATGCGTTCACCAAGTGTGGAATCACTGATCGCGCTATTGCTGCTGCGCTACGTT 1127
 QY 1293 GTCTGCTGTTACATACTGTCGCGATTTTACCGGATTAAGCAATTCATTTGTTGTTGATGA 1352
 DB 1128 GTCTAATGCTTACATCACTGATCGTCAAGTTTACCAGATTAAGCGATTCGATGATGATGA 1187
 QY 1353 GGCTTGTGCGAAATGTCGAGGTCAGCTTGTATGATCAACCTGAAGAGATTCGATACCTTGA 1412
 DB 1188 GGCTGCCAGTCTGATTTGATGGAATTTGACTCCAGCCGGAAGGCTTGTATCGTTTGA 1247
 QY 1413 AAGGAAGAGATGCGCTGGAATTTGAATTCAGCTTCGCTTGGAAAGGAGAGGATAAGC 1472
 DB 1248 GCGTCGGTTCATTCAGTTTGAATTTAGCTGAGATGCTGAGAAAGGAAAGGATGAGC 1307
 QY 1473 CAGCAAGCTCGATTTATAGAGTTCGGAAGAGCTTGTATGACTGAGACACAGCTTCA 1532
 DB 1308 GAGCAGACGCTTTAGCCGATCTTGAGCGTGTATTTGAGGTTCTGAGCCGCTGAATTTTC 1367
 QY 1533 GCCTCTCCAGTGAATAACAGAAAGGAGAGAGAAATTTGATGAGATTCGAAGGCTTAA 1592
 DB 1368 CGATCTGGAGGAGTGTGGAGATCAGAAAGGCGCGCTACAGGGGCGGCTAAGATCAA 1427
 QY 1593 ACAGAAAGAGAGAGCTCATGTTTCTTTGAGGAGGAGAGAGAGAGATATGACCTTGC 1652
 DB 1428 GGACTCGATCGAGCAGCCCAAGCTTGAATTTGGAAGCCGCGCAGCGCTCAGGACTACGC 1487
 QY 1653 AAGAGCTGCTGATTAAGATATGCGCAATTTCAAGAGTGAATCTGCAATTTGCCCACT 1712
 DB 1488 CAAGATGAGCGAGATTCAGTATGTTGCTTCTCCCGGTTGGAGAAAGCAGCTGCTGGCAGC 1547

1713	TGAAGGAAC	TCTCTCTG	AAGAGAA	TGTGAT	GTCTCA	CAGAA	AACGTTGG	CGCTG	AACACAT	1772
1548	AAGTCAG	CGGAA	CAGCACC	GA	TTTTACT	TTTGGT	GCAGAG	AAGTGA	CTGCTGAGGAGAT	1607
1773	TGCTCAG	GTGTG	TGAGCG	TTGGAC	AGGAA	TTCCAG	TGACG	AGACTTGG	CCAAAATGAGAA	1832
1608	TGCCAG	GTAGT	TAGTCG	TTG	AGTCTG	TATCCAG	TGACG	AGATGCTT	GAGGCGAGCG	1667
1833	GGAGAG	GTGAT	TGGTCT	TGCTGAT	NAG	TTGCAT	AAGCG	GGTTG	TGGGACAGAA	1892
1668	CGATAAG	TTGTTG	CGTATG	GAAC	CTGATCT	AGGTCG	ACG	GTGTGGT	TGGTCAAGAGGAGGC	1727
1893	GGTAAAT	TCAGT	TTCTCAG	GCAAT	CTTAAG	GTCAAG	GCAG	AGACTTGG	TAGGGCACACA	1952
1728	GATCAAG	GTGAT	TCGAT	CGGTG	CGCGCT	TGCGGT	CTG	TTTGTG	ATGATCCAAATCG	1787
1953	GCCAACT	TGGAT	CA	TTCTTAT	CTCTGG	ACCA	CTGGT	TTGGCAAA	AACTGAGCTCGCCAA	2012
1788	ACCGAG	CGGCTCG	TTCTCT	TTCTGG	GTCCG	ACG	GTGTGG	TAA	GA	1847
2013	GGCTCT	TGCTG	AGCAG	CTGTTG	ATGAT	GA	AAACCTCT	TAGTTCGG	ATGATGTCGA	2072
1848	GGCGCT	GGCTGA	ATTTTGG	TTTGC	ACAG	TCAAG	ATGCA	ATTTGGT	TCGCAATGATATGATGA	1907
2073	ATATATG	GAACA	CAC	CTCTGT	CTCTCG	CCCTCA	TTGGG	GCACAC	ACAGG	2132
1908	GTTCA	TGAAAA	CA	TTCTGT	GGCG	CGCTG	ATTTGGT	GGCGCT	TCGGGCTATGTTGGTTA	1967
2133	CGAGAA	GGTGG	ACA	CTTA	CTAG	GCTGTG	TGAGG	AGGCG	CACTTATTTGTGTACATCTTT	2192
1968	TGAGGA	AGGAG	GTATCT	TACT	GAATTTGGT	GCAG	CTGGCG	CTTACT	CTCTGATCTTTT	2027
2193	TGATGA	GTGGAGA	GGCTCAT	GTGCTCT	CTTCA	CACTCTG	CTCCAG	CTCTG	CTCAAGTTTGTGATGA	2252
2028	GGATGA	AGTGGAGA	GGCGAT	AGTGTG	TTCA	ATTTCTG	CTCAAG	TACTCTG	CTCAAGTACTTGTATGA	2087
2253	TGTCG	GA	TTCA	CAG	CGGCA	AGCG	AGCAG	CTGATTT	TCAGGAACTCGGTGATATCA	2312
2088	TGAGC	GGCTT	GAC	TGAT	GGTCAG	GGCG	GTACTG	TTTTCG	CAATCTGTTATTTGTGAT	2147
2313	GACAT	CAAA	CTTTGGT	GTG	TGAC	CACTCT	TCAG	GGGCTAA	CTGGGAAAGTAA	2372
2148	GACAT	CAAA	CTTAG	TTTCA	CTCAG	ATCCAA	GAATCTCAG	TGGAG	ACGATTTCTCGGAGGT	2207
2373	AGTGG	CCC-----	GG	AGCTGT	GTGAT	CGGAG	GTGAG	GAACACTTT	TCAGACACAGCT	2426
2208	GTATAC	GCAAT	GA	GGCG	CAGTGA	TGGGG	GTG	CAGCG	CA	2267
2427	CTTGAA	CAG	CGTTGA	CGA	ATTG	TGTTG	ATG	CA	CCCCCTTTCA	2486
2268	CAITAA	CAG	GTGGAT	GA	TATTTG	TTTTC	CA	CCGTTG	GA	2327
2487	AGTAG	CTCG	CTTCA	ATG	AAAG	CGTTG	CTGT	CGCG	CTGCTG	2546
2328	GATTG	CGCGA	TTCA	CTCGG	GGGCT	AG	AG	CGGTCT	CGCTGA	2387
2547	GGCAG	TCA	CTGAT	GTCTT	TGCA	CTAT	CTTTGG	CA	GAGAG	2606
2388	AGATTTG	GATGAT	CGT	GGCTG	GA	ATTG	CTTGG	TAA	TGT	2447
2607	TGCTAG	GCCTAT	AAG	GAGAT	TGGA	TGGA	AGAG	AGGTG	TG	2666
2448	TGCG	CGCG	CGCTG	AA	GAGAG	CAAT	CCA	ATCTCAG	CTG	2507
2667	TGTG	CGGTGAG	GAATCG	ATG	AA	CTCC	ACTGTTT	TAC	ATAG	2726
2508	CTTAG	CTGGAG	CGTTG	CTCAG	TGG	AGAT	ACTG	TG	CGAG	2767
2727	TGT	2729								2567
2568	TGT	2570								

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Protein - nucleic search, using frame_plus.p2n model

on: February 13, 2004, 01:41:41 ; Search time 142 Seconds
(without alignments)
2831.690 Million cell updates/sec

e: US-09-812-350-17

ect score: 4550
ence: 1 MPEKTHKTATIAHEL.....YKMRIBEDDDNEEMIED 911

ing table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

ched: 569978 seqs, 220691566 residues

1 number of hits satisfying chosen parameters: 1139956

um DB seq length: 0

um DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

and line parameters:

EL=frame+p2n.model -DEV=xlpl

cg2 1/USPTO spool_p/US09812350/runat 10022004 101516 13378/app query.fasta_1.1095

Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCHI=0

PEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi

T=45 -DCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

B=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

E=US09812350 @CGN 1.1.85 @runat 10022004 101516 13378 -NCPUS=6 -ICPU=3

MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG

TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

PEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

base :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Alt	Score	Query	Match	Length	ID	Description
1	2194	48.2	2574	4	US-08-887-534A-46	Sequence 46, Appl
2	2194	48.2	2574	4	US-09-527-431-46	Sequence 46, Appl
3	2173.5	47.8	2628	4	US-09-107-532A-883	Sequence 883, Appl
4	2162.5	47.5	1830121	4	US-09-557-884-1	Sequence 1, Appl
5	2162.5	47.5	1830121	4	US-09-643-990A-1	Sequence 1, Appl
6	2158.5	47.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
7	2158.5	47.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
8	2140	47.0	2592	4	US-09-328-352-2604	Sequence 2604, Ap
9	2133	46.9	2580	4	US-09-199-637A-280	Sequence 280, App
10	2133	46.9	2970	4	US-09-199-637A-272	Sequence 272, App
11	2133	46.9	42235	4	US-09-199-637A-1	Sequence 1, Appl
12	2132	46.9	2580	4	US-09-252-991A-4842	Sequence 4842, App

- Sequence 4897, Ap
- Sequence 1142, Ap
- Sequence 1, Appli
- Sequence 1244, Ap
- Sequence 1, Appli
- Sequence 5922, Ap
- Sequence 3, Appli
- Sequence 3, Appli
- Sequence 2055, Ap
- Sequence 1, Appli
- Sequence 123, App
- Sequence 5965, Ap
- Sequence 18, Appl
- Sequence 2033, Ap
- Sequence 168, App
- Sequence 6033, App
- Sequence 5908, App
- Sequence 684, App
- Sequence 727, App
- Sequence 661, App
- Sequence 14, Appl
- Sequence 1, Appli
- Sequence 14627, A
- Sequence 15131, A
- Sequence 3366, Ap
- Sequence 84, Appl
- Sequence 8976, App
- Sequence 2353, Ap
- Sequence 11, Appl
- Sequence 276, App

ALIGNMENTS

RESULT 1
US-08-887-534A-46
; Sequence 46, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,534A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Heien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2574 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA (genomic) (p10b30)"

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2571

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

us-09-812-350-17

635 MetSerGluTyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProProGlyTyr 654
 1900 ATCTCCGAGTTATGAGAACTCGGTCTCGTTGGTGGCTCGGGATAT 1959
 655 ValGlyHisGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysVal 674
 1960 GTCCGTTATGAGAGTGGCTACCTGACCGAGCGGTGGTCTCGTATTCGGTC 2019
 675 IleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuGlnVal 694
 2020 ATCTCTGTCGATGAAGTGAAGGAGGATCCGATCTCTCAACATCTGTTCAGGTA 2079
 695 LeuAspAspGluValLeuThrAspGlyGlnGlyValArgThrValAspPheArgSerVal 714
 2080 CTGGATGATGGCGCTCTGACTGCGGCGAGGAGGAGCGTCCGATTCGTAATACGGTC 2139
 715 IleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysVal 734
 2140 GTCAATATGACCTTAACCTCGTTCGGAT---CTGATTCAGGAACGCTTCGGTGAACATG 2196
 735 ThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGlu 754
 2197 GATTATGCGCACATGAAGAGCTGGTCTCGTGTGTGAAGCCATAACTTCGCTCGGAA 2256
 755 LeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeuArg 774
 2257 TTCATTAACGTCATGATGAGTGGTCTTCATCCGCTGGTGAACAGCACATGGC 2316
 775 LysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAla 794
 2317 TCGATGCGCAGATTCAGTTGAACGCTCTGACAAAGCTGGAAGACGTTGTTATGAA 2376
 795 LeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyr 814
 2377 ATCCATTTCTGACGAGCGCTGAACTGCTGAGCGAGAACGGTTACGATCCGGTCTAT 2436
 815 GlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMet 834
 2437 GGTGACGCTCTGAACTGCAATTCAGCAGCAGATGAAACCCGCTGGCAGCAA 2496
 835 -----ValValArgGluLeuAspGluAsnSer 844
 2497 ATACTGCTGTGTAATGGTTCCGGTAAAGTATTGCTGCTGGAAGTTAATGAAGACCG 2556
 845 ThrValTyrIle 848
 2557 ATTGTCGCGTC 2568

FILE 2

9-527-431-46
 sequence 46, Application US/09527431
 Patent No. 6485899

GENERAL INFORMATION:

APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/527,431
 FILING DATE:
 CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/887,534
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 28341/33996
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 474-6300
 TELEFAX: (312) 474-0448
 TELEX: (312) 474-6600
 INFORMATION FOR SEQ ID NO: 46:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2574 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2571
 US-09-527-431-46

Alignment Scores:

Pred. No.: 9, 01e-214 Length: 2574
 Score: 2194.00 Matches: 447
 Percent Similarity: 69.56% Conservative: 154
 Best Local Similarity: 51.74% Mismatches: 239
 Query Match: 48.22% Indels: 24
 DB: 4 Gaps: 9

US-09-812-350-17 (1-911) x US-09-527-431-46 (1-2574)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 1 ATCGCTCGGATCGCTCTACTATAAATTCAGCTGCTCTTCCGATGCCAATCACTT 60
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLysSer 40
 Db 61 GCATCTCGGCGCAGCAACCAATTTATCGAACCACTTCAATTAATGAGCGCCCTGCTGAAT 120
 QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 Db 121 CAGGAAGGGGTTCCGTTAGTCTTATTAACATCCGCT---GGCATAATGCTGGCCAG 177
 QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 Db 178 ---TTGCGCACAGATATCAATCAGGCATTAATCGTTTACCGAGGTTGAAGGTACTGCT 234
 QY 81 AspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAlaAlaGln 100
 Db 235 GGTGATGTCAGCCATCACAGGATCTGGTGGCGGCTCTTAATCTTTCGCAACAGTGGCG 294
 QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
 Db 295 CAATAACGTGTGATTAATCTCTCTCAGAACTGTTCTGTCGGCGCCTGACTGCT 354
 QY 121 ---SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 Db 355 CGCGGCACCGTGGCGGCATCTCTGAAGAGCAGCGGGCGACCCGCCCAACATTAATCA 414
 QY 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 Db 415 GCGATTGAACAAATGCGTGA---GGTGAAGCGTGAACCGATCAAGTCTGCTGAAGC 468
 QY 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
 Db 469 CAACGTGAGGCTTTGAATAAATATACCATCGACCTTACCGACGAGCCGACAGGCAAA 528
 QY 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
 Db 529 CTCGATCCGGTGATTGGTCTGTGAAGAAATTCGCGTACCATTGAGTCTGCTGCAACGT 588

198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaVal 217
 589 CGTACTAAATAACCGGTAATGATGGTGAACCGCGCTCGGTAAACATGCGCTGTT 648
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 649 GAAGTCTGGCGCAGCGTATTATCAACGGCGAAGTGGCGGGAAGGTTGAAAGGCGCGG 708
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
 709 GTACTGGCGCTGGATATGGCGCGCTGGTGGCGGAAATATCGCGGTGAGTTTGA 768
 258 GluArgLeuLysSerValLeuLysValGluAspAlaGluGlyLysValIleLeuPhe 277
 769 GAAGCTTTAAAGCGGTCTTAACGATCTTCCAAAACAGNAGGCAACGTCATCTATT 828
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 829 ATCAGCAATATACATACCATGGTGGCGCGGTAAAGCCGATGGCGCAATGGAGCCGGA 888
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317
 889 AACATGCTGAACCGCGCTGGCGGTGGTGAATGCACTCGGTAGTGGCCACGCGTT 948
 318 GluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 337
 949 GACGAATATCGCCAGTACATTAAGAAAGATGCTGGCTGGAAACGTCGTTCCAGAAAGTG 1008
 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
 1009 TTGTGTGGCGAGCCTCTGTGTGAAGATACCATTCGATTCGATTCGCGTGGCGCTGAAAGAACGT 1068
 358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 1069 TAGGAATGGACCAACCATGTCGCAATTAACCCCGCAATTTGTGACGGCGGACGTTG 1128
 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 1129 TCTCATCGTACATTCGTACCGTCAGTCGCGGATTAAGCCATCGACCTGATCGATGAA 1188
 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 1189 GCAGCATCCACATTCGTATCGATGACATCAAAACCCAGAAAGTCTCGACCGACTCGAT 1248
 418 ArgLysArgMetGlnLeuGluIleGluHisAlaLeuGluArgGluLysAspLysAla 437
 1249 CGTGTATCATCCAGCTCAAACTGGAACACAGCGGTTAATGAAGAGTCTGTGAAGCC 1308
 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGln 457
 1309 AGTAAAAAACGTCCTGGATATGCTCAACGAAGACTGAGCGACAAAGAACGTCAGTACTCC 1368
 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 1369 GAGTTAGAAAGAGTGGAAAGACGAGGACATCGCTTCTCGTACGACGACCATTA 1428
 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAla 497
 1429 GCGGAACCTGGAACAGCGAATCGCTATTGAACAGCTCGCCGTGTGGGGACCTGGCG 1488
 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGlu-----SerAlaIle 514
 1489 CGGATGTCGTAACATACGCGAATAATCCGGAACCTGGAAGAACCACTGGAAGCCGCA 1548
 515 AlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyPro 534
 1549 ACGGAGCTCGAAGGCAAAATATCGT-----CTGTTCGCTAATAAAGTACCGAC 1599
 535 GluHisIleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGln 554
 1600 CCGGAATGCTGAAGTCTGGCGGTGGACGGGATTCGGTTCTCGCATGATGGAA 1659

555 AsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGln 574
 1660 AGCGAGCGGAAAAAATCTGCTGATGGAGCAAGAACTGCACCATCGGTATTGGTACG 1719
 575 AsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArg 594
 1720 AACGAAGCGGTGATCGGTATCTAAACGCTATTCGTCGTAGCGGTGCGGGGTGGCGGAT 1779
 595 AlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlu 614
 1780 CCAATTCGCGCGATGGTTTCATCTCTCTCCGCCCAACTGGTGTGGGGAACAGAG 1839
 615 LeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAsp 634
 1840 CTTTCTAAGCGCTGGCGAATTTATGTTGATAGCGAGCGCATGGTCCGTATCGAT 1899
 635 MetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyr 654
 1900 ATGTCGAGTTTATGGAGAAACACTCGGTGTCTCGTTGGTTGGTGGCGCTCCGGGATAT 1959
 655 ValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysVal 674
 1960 GTGCGTTATGAAGAAGTGGCTACCTGACCGAAGCGGTGCTGCTCGTCCGTTATCCGTC 2019
 675 IleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuGlnVal 694
 2020 ATCTCTGCTGATGAAGTGAAGAAAGCGCATCGGATGCTTCAACATTCGTTCGAGTA 2079
 695 LeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerVal 714
 2080 CTGCGATGATGGCGCTCTGACTGACGGCAAGGAGACGGTCCGATTCGTAATACGGTC 2139
 715 IleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysVal 734
 2140 GTCATTTAGCTCTTAACCTCGTTCCGAT---CTGATTCAGGAACGCTTCGTTGAATG 2196
 735 ThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGlu 754
 2197 GATTATGCGCACATGAAGAGCTGTGCTGCGTGGTGAAGCATAACTTCGTCGCGAA 2256
 755 LeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArg 774
 2257 TTCATTAACCGTATCGATGAAGTGGTGTCTTCCATCGCTGGGTGAACACGACATGCC 2316
 775 LysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAla 794
 2317 TCGATGCGACATTCAGTTGAACGCTCTGTACAACGCTTGGAGAAGCTGTTATGAA 2376
 795 LeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyr 814
 2377 ATCCACATTCGACGAGCGCTGAAACTGCTGAGCAACGCTTGGAGAAGCTGTTATGAA 2436
 815 GlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMet 834
 2437 GGTGACGCTCTCTGAACGTCGAATTCAGACGAGATCGAAACCCGCTGGCACAGCA 2496
 835 -----ValValArgGluGluIleAspGluAsnSer 844
 2497 ATACTGCTGTTGAATTTGGTCCGGGTAAGATGATTCGCTCGGAAGTTAATGAACCGG 2556
 845 ThrValTyrIle 848
 2557 ATTGTGCGCGTC 2568

RESULT 3

US-09-107-532A-883

; Sequence 883, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

441 ArgLeuIleGluValArgGlySerLeuAspLeuArgAspLeuArgAspLeuGlnProLeuThr 460
1342 CGCTGAAAGAAATTTCAAGAAAGATTTAGCTGAATCTCTGGAAGAAAGAAATGCGATGAAG 1401
461 MetLysTyrArgLysGluLysGluArgGlySerLeuAspLeuArgAspLeuGlnProLeuThr 480
1402 ATGCAAGTGGGAAACAGAAAGAAAGAAAGTCAATCTCTGGAAGAAAGAAATGCGATGAAG 1461
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgGlySerLeuAspLeuArgAspLeuGlnProLeuThr 500
1462 GATAAGCAAGAAACAGAAATTTAGCTGAATCTCTGGAAGAAAGAAATGCGATGAAG 1521
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1522 GTATTGCGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1581
521 SerSerGluGluAsnVal--MetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
1582 GCAGAAAGACAGCGAGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1641
540 ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg 559
1642 GTCTGTCAGATTAACGGGTATCCCGTACGAACTTTGAGAGGTGACGAGAAAGAA 1701
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
1702 CTAATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1761
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
1762 CGGTGTCAGTACGGGTATTCGTTACGCGCTGATGATGATGATGATGATGATGATGATGAT 1821
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
1822 GGTCTCTTCT 1881
620 AlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMet 639
1882 GCTGAGAACTTGTGTGATCTGAGAACTTGTGTGATCTGAGAACTTGTGTGATCTGAGAACT 1941
640 GluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGlu 659
1942 GAAACACATGCTGTCT 2001
660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGlu 679
2002 GGTGTCAATTAACGAAAGTCTCAGACGAAATCTTATACAACTCTGTTGTCGATGAA 2061
680 ValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArg 699
2062 ATCGAAAGCGCATCCCGATGCTCTCAATATCTGTTACAAAGTCTGATGATGATGATGATGATGAT 2121
700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSer 719
2122 TTGACTGATTCAAAGGGCGAGTAGTAGATTTCAAAATACGGTCTCTGATCATGACCCAGC 2181
720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMet--GluVal 738
2182 AATATCGCTTACAGTACTCTGCGAAGCGCTGATGATGATGATGATGATGATGATGATGATGAT 2241
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2242 GTGGCTGAACAGTCAACAGCTATTACGTGGAACTTCAACAGCAATCTTGAACCGA 2301
759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2302 ATCGACGACACGATTTTATTTATACCACTTACCTAGCAGCGTGAAGGAAATCGTTGAC 2361
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2362 AAAATGTTGGCGCAATTCGCAACAGTCTGGAACATCAAGAAATCTCTGCTGACAAATCAGT 2421
799 AspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818

Db 2422 GATGAAGCCAAACAGTGGATTCGTAATAATGCTTATGAACCCAGCATATGCTGCAAGACCA 2481
Qy 819 IleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
Db 2482 TTGAACAGATTTATCTACTAAGGAGTAGAAACACCACTAGCTAGTAAGAAATCGTTCAGGA 2541
Qy 839 GluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArg 858
Db 2542 CATGTTATGCTTAATCAAAAGTAACGATTTACTTTACTTATGATGGCAACTTCATTTCAAA 2601
Qy 859 ValGlu 860
Db 2602 ACAGAA 2607

RESULT 4
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; FILING DATE: 25-Apr-2000
; APPLICATION NUMBER: US/09/557,884
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 5,73e-206 Length: 1830121
Score: 2162.50 Matches: 444
Percent Similarity: 69.9% Conservative: 153
Best Local Similarity: 52.05% Mismatches: 245
Query Match: 47.53% Indels: 11
DB: 4 Gaps: 8

US-09-812-350-17 (1-911) x US-09-557-884-1 (1-1830121)
Qy 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 906379 ATGACATTGAAATTTACGCAAAATTCAGAAAGCGTTAAGCGAAGCAATCACTT 906438

908581 GAAATG--AAAGCCTTAGTGTGTCAGTGTAGCCAACTTCGCCAGCAATTCATC 908637
 757 AsnArgLeuAspGluLeuValPheAspProLeuSerHisAspGlnLeuArgLysVal 776
 908638 AACCGTATTACGAAACCGGTGTTTCCATCCACTTGGTAAAGAAATATCCGTGCGATT 908697
 777 AlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAla 796
 908698 GCAAGTATCCAAATTAGACCTTAGCAGAACTGTGGAATGCAATTTGGTG 908757
 797 ValThrAspAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyValAla 816
 908758 TTTACCGATGCTTTATTAGACTTCATTGGCGAAGTGGATACGACCAATTTATGTGTGA 908817
 817 ArgProIleArgGlyTrpMetGluLysLysValValThrGluLeuSerLysMetValVal 836
 908818 CGTCCATTGAACGTGCTATTACACAGAAATTTGAAACAGCTTGGCACAAATTTCTA 908877
 837 ArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
 908878 TCTGGTGGTATTACCTGTAAGTGTGTTACCATTCAT 908916

99-643-990A-1
 Sequence 1, Application US/09643990A
 Patent No. 6528289

GENERAL INFORMATION:
 APPLICANT: Robert D. Fleischmann
 Mark D. Adams
 Owen White
 Hamilton O. Smith
 J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
 the Haemophilus influenzae Rd Genome, Fragments
 thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville,
 STATE: MD
 COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS v6.22
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A
 FILING DATE: 23-Aug-2000
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429
 FILING DATE: 1995-06-07
 APPLICATION NUMBER: 08/426,787
 FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB186P1C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-610-5790
 TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1830121 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

99-643-990A-1

Alignment Scores:
 Pred. No.: 5,73e-206 Length: 1830121
 Score: 2162.50 Matches: 444
 Percent Similarity: 69.99% Conservativeness: 153
 Best Local Similarity: 52.05% Mismatches: 245
 Query Match: 47.53% Indels: 11
 DB: 4 Gaps: 8

US-09-812-350-17 (1-911) x US-09-643-990A-1 (1-1830121)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 906379 ATGAACATTGAAAAATTTACGACAAAATTCACAGAGCGTTAAGCAGCAATTCACCTT 906438
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 Db 906439 GCTATCGGAAAAGCAATCAATTTATTGAACCTGTGCTATTTATTGACCGCACATTTTAAAT 906498
 QY 41 AspProThrGly---IlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla 59
 Db 906499 CAACAGGCGGCTCTATTGCGCCA-----ATTTGACAGCAAGTGGCGTAAATGTGCT 906552
 QY 60 GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 79
 Db 906553 ---TTATTGCGTAATGAATTAAGACACAGAACTAAATAAATTCGCCCAAGTGAATGGCAAT 906609
 QY 80 ProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAla 99
 Db 906610 GCGCGTGAATGACAACTTTCAGCCAGCTTATTAATCTACTTAATTTATTTATGGGATAAATTC 906669
 QY 100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
 Db 906670 GCACAGCAAAACCAAGATAAATTTATTCGAGCAATTTGTTTGTTCGAGCTTTAGAA 906729
 QY 120 Asp---SerGlnIleArgAspLeuAsnGluValGlyValAlaIleThrAlaArgValLys 138
 Db 906730 GAACGAGGACGATCAGCGATATTTTGAAAAGTGGCGTGCAGAAAAGAACAAATTCG 906789
 QY 139 SerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSerGlyAsp 158
 Db 906790 CAGCTATTTCAGCACTTAGAGG-----GGACAAAACGTGAACGATCAAAATGCAGAA 906843
 QY 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----Gly 176
 Db 906844 GAAACGACACAAAGCGCTTGAAAAATATACGATTGATTAAACCGCTCGTGCAAGAGTGGC 906903
 QY 177 LysLeuAspProValIleGlyArgAspGluIleArgArgValValArgIleLeuSer 196
 Db 906904 AAACCTTGATCTCTTAATTTGGCGGTGAAGAAATTCGTGAGCCATTCAGATATTACAA 906963
 QY 197 ArgArgThrLysAsnAsnProValIleGlyGluProGlyValGlyLysThrAlaVal 216
 Db 906964 CGTCTACCAAAAATAACCTGTGTAAATTTGTGAACCAAGTGTAGGAAAACCGCGATT 907023
 QY 217 ValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspVal 236
 Db 907024 GTGGAAGGCTTGGCACAGCGATCGTAAACCGCAAGTGCAGAGGCTTTGAAAAATAAA 907083
 QY 237 ArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPhe 256
 Db 907084 CGTGTGCTTTTATTAGATATGCGGCGCTGTGTTGCTGTCGAAATATTCGTGGTGAATTT 907143
 QY 257 GluGluArgLeuLysSerValLeuLysValGluAspAlaGluGlyLysValIleLeu 276
 Db 907144 GAAGAACCTTTAAAGCAGTACTCAATGAACCTTCGAAAGAGAGGTCGCTATCTC 907203
 QY 277 PheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAla 296
 Db 907204 TTTATTGACGAAATTCATCTATGTCGCGCGGTAAACCGGATGTCGATGCGATGCG 907263
 QY 297 AlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgGlyAlaThrThr 316

907264 GGTAAATTTGTTAAACCAAGTTTGGCAGGCGGAATTAACATTGCGTGGTCAACTACT 907323
 317 LeuGluGluThrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGln 336
 907324 TTAGTGAATATCGTCAATATATCGAATAAGATGCCGCACTTGAAACCGCGTTTCCAAAAA 907383
 337 ValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlu 356
 907384 GTCTTTGTGGACGAACCAAGGTAGAGAGATACCATTCGATCTTACGTGGTTTGAAGAA 907443
 357 LysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuLeuAlaGln 376
 907444 CGTTATGAATTCATCATCATCGTGGATTAATGACCCAGCAATTTGCGTCGACGAACG 907503
 377 LeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAsp 396
 907504 CTTCACATCGTTATATTCGATCGTCAGTTACCATTAAGCAATTTGATTTGATCGAT 907563
 397 GluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeu 416
 907564 GAAGCAGCGCTAGCATTCGTATGTAATAGATCTTAACCTGACCGCTTGATCGCTTT 907623
 417 GluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgLysAspLys 436
 907624 GAACGTCGTATTATCCCAATTAATTTGAATTCGAACAACAGCGTTTCAAAAAGAAAGACGAA 907683
 437 AlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeu 456
 907684 GCAAGTCGCAACCTTTAGAAATGTAGAGAAAGATTTGGCTGAAAGAAACGCTGAATAC 907743
 457 GlnProLeuThrMetLysTyrArgLysGluArgIleAspGluIleArgArgLeu 476
 907744 GCCGAACCTTCAAGAGATGTAATCTGAAAGCAACGCTTTCTGGCTCTCAACATATT 907803
 477 LysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeu 496
 907804 AAACAAGAGTTAGTACTGCAAAACCCGAACCTAGCAACAGCTCGTCGCGGGTGATTTA 907863
 497 AlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGln 516
 907864 CGCAATATGTCAGTTGCAATATGCGCGCATCCCTGATCTTGAAAGCAACACTTGAACAA 907923
 517 LeuGluGlyThrSerSerGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
 907924 GCTGAAACCGACGAAGAAAGAAATGACGCTTTACGCTATCGCTCACAGTGAAGAA 907983
 537 IleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
 907984 ATCCGAGAAGTGTCTTTAAAGCCACAGGCAATTCCTGTATCAAAAATGATGAAGCGAG 908043
 557 LysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
 908044 AAAGAAAACCTTTGCGTAGGAGATGAATCACTAATAACGAGTGAATTTGTCGAAGAA 908103
 577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
 908104 GCGGTTGATCGGTAGCAACAGGATTCGTCGTAGTCGTGAGGCTCTTCCGATCTCTAAT 908163
 597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
 908164 CGCCCAATGTTCTTTCTTTGTTCTTGGGCCCAACAGGTTTGGGAAAACAGAGCTTTGC 908223
 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSer 636
 908224 AAACTTTGGCTAAATCTTTGTTGATGTAAGATGCGATGGTGGTATTGATATGTCA 908283
 637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGly 656
 908284 GAGTTTATGAAAAACACACAGTGTCTCGTTAGTTGGTGGCTCCAGGCTATGTCGC 908343
 657 HisGluGluGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeu 676
 908344 TATGAAGAGCGGTTTATTAACTGAAGCTGTTGTCGTGCTGCTCATATTTCAGTGTCTTA 908403

QY 677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 696
 Db 908404 TTAGATGAAGTTGAAAAAGCACACACAGATGATTCAATATCTTATTACAGTGTGGAT 908463
 QY 697 AspGlyValGluThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIle 716
 Db 908464 GATGGTCGTTTAACTGATGGTCAAGTGTGACTGTGGACTTCCGTAACACTGTGGTTAT 908523
 QY 717 MetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMet 736
 Db 908524 ATGACCTCTAACTTGGGTTCTGAT--TTAATCCAAGGTAAATAAAGACGAAAGCTATAGC 908580
 QY 737 GluValAlaAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeu 756
 Db 908581 GAAATG--AAAGCCCTTAGTGTGTCAGTGTGTAAGCCCAACATTTCCGCCCAAGATTTCATC 908637
 QY 757 AsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysVal 776
 Db 908638 AACCGTATTGAGAAACCGTGGTATTCCATCCATCTTGTAAGAAAATATCCGTGCGGATT 908697
 QY 777 AlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAla 796
 Db 908698 GCAAGTATCCCAATTAGAACGCTTAGCAAAACGATGGAACCTCGTGGATACGAATTGGTG 908757
 QY 797 ValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAla 816
 Db 908758 TTTACCGATGCTTTATTAGACTTCATTGGCGAAGTGGATACGACCAATTTATGGTGCA 908817
 QY 817 ArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValVal 836
 Db 908818 CGTCATTTGAACGCTGCTATTTCACACAGAAATTTGAAACACAGCTTGGCACAAATTTCTA 908877
 QY 837 ArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
 Db 908878 TCTGGTGGCTTATTACCTGCTAAGGTTGTTACCATTTGAT 908916

RESULT 6

US-09-103-840A-2/c
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103.840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 4403765
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURES:
 ; OTHER INFORMATION: CDC 1551
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence
 ; OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Alignment Scores:
 Pred. No.: 6.03e-205 Length: 4403765
 Score: 2158.50 Matches: 447
 Percent Similarity: 69.66% Conservative: 150
 Best Local Similarity: 52.16% Mismatches: 227
 Query Match: 47.44% Indels: 33
 Gaps: 9

US-09-812-350-17 (1-911) x US-09-103-840A-2 (1-4403765)

2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
462071 AACCCG-----ACGACCAAGACGACGCGCGCTTAACCGCGGCTTACAGCGCGCT 462021
22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisIleuAlaGlyAlaLeuIleSerAsp 41
462020 TCGACCGCGCGCAATCCGAGATCCGCGCTCACCTGCTGTAGTGGCGCTGCTGACCCAA 461961
42 ProThrGlyIlePheProGlnAlaIleSerAlaGlyGlyGluAsnAlaAlaGlnSer 61
461960 AACGACGATATGCGCGCACCGCTACTGAGGCTGTGCTGTCGACCGCGCCACCGTCCGC 461901
62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
461900 GCCGAA-----ACCCAGCGCTGCTGACCGCTTTCGCGAGCGGCTGAGGCGCAGCACG 461847
82 AspileProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGlnLys 101
461846 CAGCGCGAGCTGCTCCCGGAGTGTGTAGCGCGATCACCGCGCGAGCTGCGCCACC 461787
102 SerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu---Asp 120
461786 GAGTGGACGAGAGTACTCTCCACCGAGCAGCTGATGCTGCGGCTGGCCACCGGTGAC 461727
121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
461726 TCCGACGTCGCAAGCTTTGACCGCGCCACGCGCGCTCGCGCAGCGGCTGCGGGAGCG 461667
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
461666 TTGCTCAAGTGGCGCGC-----AGCGCCCGGTGTACCGCCCGAACCAGCGGCGAC 461613
161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLysLeu 178
461612 TATCAGCGCTGCAGAGTACTTCCACCGACCTGACCGCGCGCGCGCGGAGGCAACTC 461553
179 AspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgArg 198
461552 GACCGGTTCATCGCGCGCGCAACAGAGATCCCGCGGTGTCGAGGTGCTGTCCCGTCGC 461493
199 ThrLysAsnProValLeuIleGlyLysProGlyValGlyLysThrAlaValAlaGlu 218
461492 ACCAAGAACACCGGTGCTGATCGGTGAGCGCGCGGTTCGCAAGACCGCGATCGTGAG 461433
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
461432 GGCCTGGCGCAGCGCATCGTCCGCGCGACGTGCGCGAGAGCTTTCGCGCAAGACCATC 461373
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
461372 GTCCGCTCGATCTCGGCTCGATGCTGCGCGCGCTCCCAATACCGCGCGAATTCGAGGA 461313
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
461312 CGGCTCAAGCGCTCTCTCGACATCAAGAACTCGCGCGGCCCAATCATCATCTTCATC 461253
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThr---GluGlySerMetAspAlaAla 297
461252 CACGAGCTGCACACCATCTCGCGCGCGCGCCACCGCGAGGGGGCGATGCGACGCGCG 461193
298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
461192 AACATGATCAAGCGATCTGCGCGCGCGCGGAGTTACGGCTGCGCGCGGCGACCGCTG 461133
318 GluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 337
461132 GACGAATACCGAAGACATCGAGAGGACCGCGCGCTCGAGCGCGCTTTCACACAGGTG 461073
338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
461072 TACGTGCGGCGCGCTGCGTGGAGGACACCATCGCATCTCTGCGCGCGCTCAAGACCGC 461013
358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377

Db 461012 TACGAGGTGCACCCAGCGGTGCGCATCCGACTCGCGCTGGTGCAGCTGCCACTTGG 460953
QY 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
Db 460952 AGCGACCGGTATATACCCCGCTTCTTCCGCCGCAAGGCCATCGACCTGTGTGACGAG 460893
QY 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
Db 460892 GCGGCCAGCGCGTGGGATGAGATCGACTCGCGCGCTCGAGATCGACGAGTCCAG 460833
QY 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
Db 460832 CGGCTGCTGCGCGCTGGAGATCGAAGAGATGCGCTGTCCAAAGAAAGACGAGGCG 460773
QY 438 SerLysAlaArgLeuIleLeuValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
Db 460772 TCGCGGAGCGGTGGCCAAAGCTGCGCTCGGAGCTGGCGGACACAGAAAGAGAGTTGGCC 460713
QY 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
Db 460712 GAGCTCACCCAGCTGGCAGAACCGAAGAACGCGATCGAAATCGTCCGCGACCTCAAG 460653
QY 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAla 497
Db 460652 GAGCAGCTGGAAGCCCTGCGCGGGAATCCGAGCGGCGGCGAACGCGACCTGCGCC 460593
QY 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIle---AlaGln 516
Db 460592 AAGCGCGCGAGCTGCGCTACGAGCGATCCCGAGGTGGAGAAAGCTCGACCGCGCG 460533
QY 517 LeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
Db 460532 TTGCGCAGCGCGCAGCGCGGAGCAGAGTGTGTCACAGGAGGAGTCTCGTCCCGACGAC 460473
QY 537 IleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
Db 460472 ATCGCGCAGCTGCTGCGCGGTGAGCCGCGATCCCGCGCGGTCTCGGAGGCGAG 460413
QY 557 LysGluArgLeuIleLysLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
Db 460412 ACCGCCACAGCTCTGCGCATGGAAGACGAGCTGGCGAAGCGGTCTATCGGGCAGAGGCC 460353
QY 577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
Db 460352 GCGGTTACCGCAGTCTCTGATCGGTGCGCGCGCAGCGCGCGGTGTCTCGACCCCAAC 460293
QY 597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
Db 460292 CGGCCACCGCGCGGTTCATGTTCTCGCGCGCAGCGGTGTCGCAAGACCGAGTGGCC 460233
QY 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSer 636
Db 460232 AAGCGCTGCGCGACTTCTCTGTCGACGACGAGCGGCGATGGTCCGATCGACATGAGC 460173
QY 637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGly 656
Db 460172 GAGTACGCGAGAACACACCGTGTGCTGTGATCGCGCGCGCGCGCTATGTGGGA 460113
QY 657 HisGluGluGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeu 676
Db 460112 TACGAGCGCGCGGTGAGTACCGAGCGGTGGCGCGGTTCACCGGTGGTGGTGTG 460053
QY 677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 696
Db 460052 TTCGACGAGATCGAAGAGGCGCACCGGACGCTGTTCGAGCTGCTGTCAGGTCTCTGAC 459993
QY 697 AspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIle 716
Db 459992 GAGGCGCGCTCACCGACGCGGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 459933
QY 717 MetThrSerAsnLeuGly-----AlaGluHisLeuLeuAlaGlyLeuThrGly 732

```

459932 CTGACGTCCAACCTGGGTGGGTGGCAGCGCCGAGCAGGTGCTGCCCGCGG----- 459882
733 LysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArg 752
459881 -----GTGCGCGCTACGTTCAAG 459864
753 ProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGln 772
459863 CCGGATTATCAACCGGCTCGACGAGTGTCTATCTTTGAGGGTCTCAACCCCGAAGAG 459804
773 LeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGly 792
459803 CTGCTGCGCATCTCGACATCCAGCTGCGCGCAGCTGGCAAGCGGTGCGCAGCGCGG 459744
793 ValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspPro 812
459743 CTCACGTGTCAGTCTCTGTCGCGGCAAGCGTGTGCGCAGCGCGGATTTCGACCCG 459684
813 ValTyrGlyAlaArgProIleArgAlaGlyTyrMetGluLysLysValValThrGluLeuSer 832
459683 GTGTAGCGGCGCGCGTGTGCGCGGCTGGTGCAGCAGCGCATCGTGACCACTGGCC 459624
833 LysMetValValArgGluGluLeuLeuAspGluAsnSerThrValTyrIleAsp 849
459623 AAGATGCTGTGGCGCGCAGGTGCAGCAGCGGATACCGTGGCGGTCAAC 459573

LT 7
9-103-840A-1/c
quence 1, Application US/09103840A
tent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
PPLICANT: WHITE, Owen R.
PPLICANT: FRASER, Claire M.
PPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
ILE OF INVENTION: TUBERCULOSIS
ILE REFERENCE: 24366-20007.00
URRENT APPLICATION NUMBER: US/09/103,840A
URRENT FILING DATE: 1998-06-24
UMBER OF SEQ ID NOS: 2
FTWARE: PatentIn Ver. 2.1
Q ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
9-103-840A-1

ment Scores:
. No.: Length: 4411529
e.: 2158.50 Matches: 447
ent Similarity: 69.66% Conservative: 150
Local Similarity: 52.16% Mismatches: 227
Y Match: 47.44% Indels: 33
Gaps: 9

9-812-350-17 (1-911) x US-09-103-840A-1 (1-4411529)
2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
461988 AACCCG-----ACGACCAAGACGCGCGCGCTAACCGCGGTTCACGCGGCT 461938
22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAsp 41
461937 TCGACCGCGCAATCCCGAGATCCGCGCGCTCACCTCTGATGCGCGTCTGCTGCA 461878
42 ProThrGlyIlePheProGlnAlaIleSerAlaGlyGluAsnAlaAlaGlnSer 61
461877 AACGACGGTATGCGCCACCGTACTGAGGCTGTGAGTGTGAGCCCGCACCGTCCGC 461818
62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
|||||
|||||
|||||
|||||

```

438 SerLysAlaArgLeuLeuValArgLysGluLeuAspLeuArgAspLysLeuGln 457
460689 TCGCGGAGCGGTGGCCAAAGTCCGAGTCCGCGACCGACGAGAGAGAGTGGCC 460630
458 ProLeuThrMetLysTyrArgLysGluLysGluArgLysGluLeuArgLys 477
460629 GAGCTCACACCCCGTGGCAGAACGAGAGAACGCGATCGAAATCGTCCGCGACTCAAG 460570
478 GlnLysArgGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
460569 GAGCAGCTGGAAGCCCTCGCGGGGGAATCCGAGCGGCGGAGACGCGGACCTGGCC 460510
498 ArgAlaAlaAspLeuArgTyrGlyAlaLeuGlnValGluSerAlaLeu 516
460509 AAGGCGCGGAGCTCGCTACGAGCGCATCCGAGGTGGAGAGAGCTCGACGCGCG 460450
517 LeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
460449 TTGCGCAGCGCGAGCGCGGAGCAGGTGATGCTCAAGGAGGAGGTGCTCCCGCAGC 460390
537 IleAlaGluValValSerArgTyrThrGlyLeuProValThrArgLeuGlyGlnAsnGlu 556
460389 ATCGCGGAGCTGTGCTCGCGTGGACCGGATCCGCGCGGTGCTGCTGGAAGCGAG 460330
557 LysGluArgLeuLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
460329 ACCGCCAAGCTGCTCGCATGAGACGAGCTGGGCAAGCGGTGTCATCGGCGAGAGGCC 460270
577 AlaValAsnAlaValSerGluAlaLeuLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
460269 GCGGTACCGCAGTCTCTGATGCGTGGCGGCGACCGCGCGGCGGTGTCGACCCCAAC 460210
537 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
460209 CGGCCACCGCGCGGTTCATGTTCTCGCGCGACCGGTGTCGCGCAAGACCGACCTGCCC 460150
617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgLysMetSer 636
460149 AAGCGCTGGCGCACTTCTGTTGACGACGAGCGGCGGTGTCGCGCATCGCATGAGC 460090
637 GluTyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProGlyTyrValGly 656
460089 GAGTACGCGGAGAGCACACCGTGTGCTGCTGATCGCGCGCGCGCGGTGTCGCGG 460030
657 HisGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValLeuLeu 676
460029 TACGAGCGCGCGGTGCTGCTGACCGGCGGTGCGCGCGGTGCTGCTGCTGCTGCTG 459970
677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 696
459969 TTCGACGAGATCGAAGAGCGCACCGCGACGTGTCGACGCTGCTGCTGCTGCTGCTG 459910
697 AspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValLeuLeu 716
459909 GAGCGCGCGCTCACCGACGCGCACCGCGCGACGTGCTGCTGCTGCTGCTGCTGCTG 459850
717 MetThrSerAsnLeuGly-----AlaGluHisLeuLeuAlaGlyLeuThrGly 732
459849 CTGACGTCACACCTGGGGTGGGTGGCGGCGCGCGCGGTGCTGCTGCTGCTGCTG 459799
733 LysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysPheArg 752
459798 -----GTGCGCGCTTACGTTCAAG 459781
753 ProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGln 772
459780 CCGGAGTTCATACACCGCTCGACGAGCTGCTCATCTTTGAGGGTCTCAACCCCGAAG 459721
773 LeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgLys 792
459720 CTGTGCGCATCGTGCATCATCGTGGCGGAGCTGGCGCAAGCGGTGGCGCGCGG 459661

793 ValAlaLeuAlaValThrAspAlaLeuAspTyrIleLeuAlaGluSerTyrAspPro 812
459660 CTCAGCTGCAGGTCTCGTGGCGGCAAGCGTGGTGGCGAGCGGATTCGACCGG 459601
813 valTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSer 832
459600 GTGTACGGGCGCGCGCTTGGCGCGGTGGTGGCGAGCGCATCGTGCACGCTGGCC 459541
833 LysMetValValArgGluGluLeuAspGluAsnSerThrValTyrIleAsp 849
459540 AAGATGCTGTGGCGCGCGGAGTGGCGAGCGGATACCGTGGCGGTCAAC 459490

RESULT 8

US-09-328-352-2604
; Sequence 2604, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2604
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2604

Alignment Scores:
Pred. No.: 2,966-208 Length: 2592
Score: 2140.00 Matches: 424
Percent Similarity: 70.35% Conservative: 181
Best Local Similarity: 49.30% Mismatches: 245
Query Match: 47.03% Indels: 10
Gaps: 6

US-09-812-350-17 (1-911) x US-09-328-352-2604 (1-2592)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 13 ATGCGATTGTGAATAATTACGAAACCGCTTGGACGAGCCCTCTCAGATGCTCAATCCTTA 72
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40
Db 73 GCGATGGTAAAGACCATACAGCTATAGCAGGTATTATTTGAGTACTTTATTGCAA 132
QY 41 AspProThrGlyLeuPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Db 133 GAGCGCTCCAATATT-----AGTTGTTGGCAACGAGCGGTGACCG-----TTACCT 180
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
Db 181 GAACCTTAAACAAAGCTAGACGAGCTTTAAAGATGCTCCAACTATTGCTAACCCGAGC 240
QY 81 AspAspIleProAlaSerSerSerLeuLeuLysValIleArgArgAlaGlnAlaAlaGln 100
Db 241 GCGCATGTCAATTTAAACCCAGACGAGTTAAAGCACTCAACTTGGCAGATCGATACGCG 300
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp 120
Db 301 CAAAAGCTGGCGATGAGTGTGTTGCTCACTGCTGGGTGTTTATGGGCTTGGCGAAACT 360
QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
Db 361 GGTGAACAAATAATTTAAGTCCGCTAGGTGTAACTCCCGACAGCTTACGCAAGTA 420
QY 141 ValGluLysLeuArgGlyLysGlyLysValGluSerAlaSerGlyAspThrAsn 160
Db 421 ATTGAATAATTCAGGAGTAGTGAC-----AAAGTCATGAGTAATAATACGAGACCAA 474
QY 161 PheGlnAlaLeuLysThrTyrClyArgAspLeuValGluGlnAla-----GlyLysLeu 178

475 CGTGACTCACTTAATAAATATACGATTGATTAAACCGCGGCTTTATCGGGAACCTT 534
179 AspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgArg 198
535 GATCCAGTGATTGGCGCGTATGATGAGATCGCGGTACCATTCAGTCTTGTACGCGGT 594
199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
595 ACTAAATAATACCCAGTACTCATTTGGTGAACCTGGGTAGGTAAACCGCTATTGTTGAA 654
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
655 GGTTCGCACAACTGATTGTCAATGGTGAAGTACCAAGAGCTTAAGAAATAAACGCTGT 714
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
715 TTATCGTTAGATTAGGTTCATTACTTTCGAGTCTCAAGTATCGTGTGATTTGAAGAG 774
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
775 CGTTAAAGCTGTTTAAAGATTGGCGAAACACCAAGCGCAATCATCTTATTCAAT 834
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
835 GACGAGTTACATACACTCGTTGGTGTCTGTTAAAGTGCACGCGCGATGGATGCAGGTAAT 894
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
895 ATGTAAACACCTGGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 954
319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyr 338
955 GAATATGCCAATACATTGAAAGATGCAGCTTGGAGCGTCTGTTCCAAAGTGTCTG 1014
339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
1015 GTCGATGAACCAAGTGTAGAGATACCATTCGATTCGATTCGATTCGATTCGATTCGATTC 1074
359 GluGlyHisHisGlyValArgIleGluAspArgAlaLeuIleAlaGlnLeuSer 378
1075 GCGACTCACCATCGCGTACGATTTTGAATTCAGCTCAGCATTTATGTCGCGGAAATGTCT 1134
379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1135 CATGTTATATACAGCCGTCAATTACAGACAGCGGATGACCTGATTTGATGAGGCC 1194
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
1195 GCTTCTCGTATTAAGATGGAATTCGATTTAAGCCAGACGACTTGTATAAATTTGATCGC 1254
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1255 GGTTCATCCAGTTGAAATGCAATTCGAACCGGTGAAAGAAAGATGAAGACGCGAGCAGT 1314
439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458
1315 AAGCCGAAGTACTCATCTTGAAACACAGATCGCTGAAGTCGAGAAAGATACAAGAT 1374
459 LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGln 478
1375 CTGAAGAAGTGTGGAAGCTGAGAAACACATGCTAGAGGCACTAAACAAAGCTCAGGTT 1434
479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArg 498
1435 GAATTAAGACACGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1494
499 AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
1495 GCACGAGTTTGCATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1554
519 ---GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537

1555 GTTGTGTAAGAAAAAGAGAGCCAAACTCATTCGTACAAAAGTAAGTGAATAATGAAATT 1614
538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
1615 GCCGAAGTCGTTAGTGTCTCAACAGATATTCGGTTGCTTAAATATGATGCAAGCGAGCGT 1674
558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
1675 GAAAAACTCTTCATATGAAGAGTTCCTTGCATGCCGTGTTGTAGGGCAAGATGAAGCA 1734
578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
1735 GTTCGTTCGGTATCGAATGCTGTTCCCGTTCCTGCTGCTGTTCTGACCCGATCGT 1794
598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
1795 CCTAGCGGATCATCTCTGTTTGTAGGACCAACAGGTGTTGGTAAACTGAGTTGACTAAA 1854
618 AlaLeuAlaGluLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
1855 GCTTAGCTAACTCTTCTTGCAGTGCATGATGCCATGATTCGTATCGATGATGATGAA 1914
638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis 657
1915 TTTCATGGAGAAACATTCGTACAGCGTTCGTGTCGCTCCGCTTATGATGATGTTAC 1974
658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
1975 GAAGAGGGCGGTGTTTAAACCAAGCTGTTCCCGTAAACCATATAGCGTAGTGTGTT 2034
678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 697
2035 GATGAGTTGAAAGAGCGCATCCAGATGCTTCAATATCTTGTACAAAGTGTAGACAT 2094
698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
2095 GCAAGCTTAAACGACTCAACAGTCTGTAGTGGACTTTAAAAACACCGTTTATGATG 2154
718 ThrSerAsnLeuGlyAlaGluHisLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
2155 ACATCGAATTCGGTTCACAGATGTCGTAACCTTGGTGAAGGTGCAACTGATGATGAA 2214
738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeu 757
2215 GTG---CGTACTTATGTAATGAATCGGTAGTACGATTCCTCCGTCAGAAATTTATTAA 2271
758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
2272 CGGATTCGATGAGTGTGATTTTCCATTCACCTCAAAAAGCACAGATTCGTGGCATTGCC 2331
778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
2332 GATATTCAGTTGGACCGCTTACGCTCAGACTTGTGTGATCGTATATGATGATGATGAT 2391
798 ThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyValArg 817
2392 GATGACAGTGCATTTGACTTATGATTCACGCTGTTGATTCCTCTATATCGAGCGCGT 2451
818 ProIleArgArgTyrMetGluLysValValThrGluLeuSerLysMetValValArg 837
2452 CCATTGAAACGTGCAATTCACACAGGTTGAAATACACTAGCTCAAAAATCTTGTCA 2511
838 GluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyr 857
2512 GGTGACTTGTTCGGGTGATACCATTTTAAAGGTGAAATGGTCACTTAGTGT 2571

RESULT 9

US-09-199-637A-280
; Sequence 280, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.

APPLICANT: Rahme, Laurence G.
 APPLICANT: Mahajan-Miklos, Shalina
 APPLICANT: Tan, Man-Wah
 APPLICANT: Cao, Hui
 APPLICANT: Drenkard, Eliana
 APPLICANT: Tsongalis, John
 TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 FILE REFERENCE: 00786/361002
 CURRENT APPLICATION NUMBER: US/09/199,637A
 CURRENT FILING DATE: 1998-11-25
 PRIOR APPLICATION NUMBER: 60/066,517
 PRIOR FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 280
 LENGTH: 2580
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 09-199-637A-280
 Alignment Scores:
 d. No.: 1,52e-207 Length: 2580
 2133.00 Matches: 427
 69.95% Conservative: 176
 49.54% Mismatches: 245
 46.88% Indels: 14
 4 Gaps: 6
 09-812-350-17 (1-911) x US-09-199-637A-280 (1-2580)
 1 MetAsnProGluLysPheThrHisLysThrAnGluThrIleAlaThrAlaHisGluLeu 20
 16 ATGCGAATAGACCGTTTACCAAGCAAGCTGCACTGGCGCTCTCCGAGCGCCAGTCCCTG 75
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
 76 GCGGTGGCCATGACCATCCGCCATCCAGCGCGGTGACCTGCTTTCGCGCTCTCGAG 135
 41 AspProThrGlyLeuPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 136 CAGCAAGCGGTTCGATCAAGCCCTGCTGATGACAGGTGCGTTCGATATCGCGCCCTG 195
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
 196 CGCAGCGGC-----CTCAACAAGAACTCGACGCGCTGCGCAAGATCCAGAGCCGAC 249
 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAlaGln 100
 250 GCGCAGCGTAACCTGTCAGGATCTCGACGCTGCTCAACAGCTGACCGCTGCGCTGCC 309
 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuLeuMetGlyLeuLeu---Glu 119
 310 CAGCAGAGAGCGGACCACTTCTCCAGCAGGTGTTATGCTGGCGCGATGAGCAGAG 369
 120 AspSerGlnIleArgAspLeuAsnGlnValGlyValAlaThrAlaArgValLysSer 139
 370 AACACCGCTCGGCAAGCTGCTGCTCGCGCAGGCGGTGTCGCGCAAGCGCGCTCGAAGT 429
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 430 GCGGTGGCAACCTGCGTGGC-----GGCGAAGCGGTGACAGCAGCCCGACGTCGAGAG 483
 160 AsnPheGlnAlaLeuLysThrTyGlyArgAspLeuValGluGlnAla-----GlyLys 177
 484 TCGCGCCAGGCGCTGGACAAGTACACCGCTGCATGACCAAGCGCGCGCGGAGGCAAG 543
 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValAlaGlyLeuSerArg 197
 544 CTCGACCGGTGATCGGTGCGCAGCAGATCCCGCGACCATCCAGGTCTCTGACGCG 603
 198 ArgThrLysAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217

Db 604 CGGACCAAGAAACCAACCGGTGCTGATCGGCGAAACCCGCGCTCGCAAGACCGCCATCGTC 663
 Qy 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 Db 664 GAGGCTGCGCCAGCGCATCATCAACGCGAAGTCCGAGCGCTCAGGACCAAGCGC 723
 Qy 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGluPheGlu 257
 Db 724 CTGCTGCGCCCTGGACATGGGGCGCTGATCGCGGTGCCAAGTTCGCGCGAGTTCGAG 783
 Qy 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 Db 784 GAACGCTGAAGCGGTCTCTCAACGAACCTGGCAAGCAGCAAGCGCGGTCTATCTGTTT 843
 Qy 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 Db 844 ATCGACGACTGCACCATGTTGTCGCGCGCGCAAGCGGAGGTGCCATGGACCGCGC 903
 Qy 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317
 Db 904 AACATGCTCAAGCGGTCTGCGCGCGCGAGCTGCTGCTGCTGCTACTACCTC 963
 Qy 318 GluGluTyArgLysTyValGluLysAspAlaAlaPheGluArgPheGlnVal 337
 Db 964 GACGATATCGCCAGTACATCGAAGAGTTCGCGCTGAGCGCGCTTCCAGAGGTG 1023
 Qy 338 TyValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 Db 1024 CTGCTGGACGACCGAGGAGACCATCGCTCTCTGCTGCTCAAGGAACGC 1083
 Qy 358 TyGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 Db 1084 TATGAAGTGCACCGCGGTGAGCATCCGCGCGCGCATCATCGCGCGCGCAAGCTG 1143
 Qy 378 SerAlaArgTyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 Db 1144 TCGACCGCTTACATACCATCCGATCGGCAACTGCGCGCAAGCGCATCGACTGATGACGAG 1203
 Qy 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 Db 1204 GCGCGCAGCGCATCGCATGGAGTCTCAAGCGCGAGAACTGATCTGCTCTGAC 1263
 Qy 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 Db 1264 CGTCGCTGATCCAGTGAAGTTCGAGCGCGAGCGCTGAGCAAGCAAGCAGCAGCGC 1323
 Qy 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 Db 1324 ACCAGAAAGCGCTGCGCAAGCTGGAGGAGGATATCGTCAAGCTCGAGCGCGAATACGCC 1383
 Qy 458 ProLeuThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 Db 1384 GACCTCGAGGAGATCTGGAAGTCCGAGAGCGCGAGGTGCGAGGCTCGCGCGCATCCAG 1443
 Qy 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyAspLeuAla 497
 Db 1444 CAGAAATCGACGACCAAGCAGGAGATGAGGCGCGCGCGCGCAAGCGCGACCTCGAG 1503
 Qy 498 ArgAlaAlaAspLeuArgTyGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 Db 1504 AGCATGCGCGCATCCAGTACAGACCATCCGACCTCGGAACCGCGCTCGAGATGTC 1563
 Qy 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 Db 1564 GACCAAGCGCAAGCAGCAAGCAAGTCTGTCGCAACAGGTGACCGCAGCAGGAAATC 1623
 Qy 538 AlaGluValValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 Db 1624 GCCGAAGTGTGTTCCAGTGGACCGGTATCCGCTGTCGAGATGCTCGAGCGCGAGCGC 1683
 Qy 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 Db 1684 GAGAGCTGCTGCGCATCGAGCAGGAGCTGCTATCGCGAGTGTATCGCGCAGGAGCGCG 1743

```

; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-272

Alignment Scores:
Pred. No.: 1.91e-207 Length: 2970
Score: 2133.00 Matches: 440
Percent Similarity: 68.90% Conservative: 178
Best Local Similarity: 49.05% Mismatches: 257
Query Match: 46.88% Indels: 24
DB: 4 Gaps: 8

US-09-812-350-17 (1-911) x US-09-199-637A-272 (1-2970)
QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrAlaThrAlaHisGluLeu 20
Ddb 2949 ATGCGATAGACCGTTTGACCGACGAGTCGCAACTGGCGCTCTCGAGCCCACTCCCTG 2890
QY 21 AlavalAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Ddb 2889 GCCGTGGCCATGACCATCCGCGCATCGAGCCGGTGCACCTGCTTCGCGCTCTCGAG 2830
QY 41 AspProThrGlyLePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Ddb 2829 CAGCAGGCGGTTCGATCAAGCCCTGCTGATGCGAGTTCGGCTTCGATATCGCGCCCTG 2770
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
Ddb 2769 CGCAGCGGC-----CTCAACAAGAACTCGACGCGCTGCGAAGATCCAGAGCCCGACC 2716
QY 81 AspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAlaAlaGln 100
Ddb 2715 GCGCAGCTGAACCTGCCAGGATCTCCAGCGCTGCTCAACACGAGGCTACCGCTCGCC 2656
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu---Glu 119
Ddb 2655 CAGCAGAGGCGGACCGATTTCATCTCCAGCGAGCTGTTATGTCGGCGCGATGGACGAG 2596
QY 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
Ddb 2595 AACACAGAGCTCGGCAAGCTGCTGCTCGCGCAGGCGCTGTCGCGCAAGGCGCTCGAGAAT 2536
QY 140 GluValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSerGlyAspThr 159
Ddb 2535 GCGCTGGCCAACTGCTGGC-----GGCGAAGCGTGAACGACCGAAGTCGAGGAG 2482
QY 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
Ddb 2481 TCGCGCGCAGCGCTCGACAAGTACACCGTCGACATGACCAAGCGCGCGAGGAAGCAAG 2422
QY 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValargIleLeuSerArg 197
Ddb 2421 CTCGACCCCGTGTATCGGTTCGCGACGACGAGATCCGCGGACCATTCACAGTCTCGACGCG 2362
QY 198 ArgThrLysAsnAsnProValIleGlyGluProGlyValGlyLysThrAlaValVal 217
Ddb 2361 CGGACCAAGAAACAACCCCGTGTGTATCGGCGAACCCCGCTCGGCAAGACCGCCATCTGTC 2302
QY 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
Ddb 2301 GAGGGCTTGCCACGCGATCATCAAGGGCAAGTGCAGCGCTCAAGGACGAGCGC 2242

```

2481	TCGCGCCAGGCGCTGGACACAGTACACCGTCGACATGACCAAGCGCGCCGAGGAGGACGAG	2422
db		
2481	TCGCGCCAGGCGCTGGACACAGTACACCGTCGACATGACCAAGCGCGCCGAGGAGGACGAG	2422
db		
178	LeuAspProValIleGlyArgAspIuclulIleArgArgValValArgIleLeuSerArg	197
2421	CTCGACCCGGTGATCGGTGCGAGCAGGAGATCGCCGGACCATCCAGGTCCTCGACGCG	2362
db		
198	ArgThrIysAsnAsnProValIleuIleGlyGluProGlyValGlyIysThrAlaValVal	217
2361	CGGACCAAGAACCAACCCGGTGCTGATCGGCGAACCCGCGTCGGCAAGACCGCCATCGTC	2302
db		
218	GluGlyIleuAlaGlnArgIleValIleGlyAspValProAsnSerIeuThrAspValArg	237
2301	GAGGGCTGGCCAGCGATCATCAACGGGCAAGTGGCGACGGCTCAAGGACAGCGC	2242
db		

598	ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys	617
1161	CGAGCGGCTCGTCTCTCTCGCGCCGACCGGGTGGCAAGACCGAGTGTGCAAG	1102
618	AlaLeuAlaGluLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu	637
1101	GGCTGGCGCGAGTCTCTTCGATACCGAGGAGCGCTGTCGGGATAGATATGTCGAG	1042
638	TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis	657
1041	TTATGAGGAAACATCGGTGGCCGCTGATCGCGCCCTCCGGGCTACGTCGCTTC	982
658	GluGluGlyGlyGlnLeuThrGluAlaValArgArgGlyProTyrCysValIleLeuPhe	677
981	GAGGAAGCGCGTACTTGACCGAGCGGATCCGCGCAAGCCCTACTCGGTGGTCTGCTG	922
678	AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp	697
921	GACGAGGTGGAGAGGCCCATCCGGATGTATTCAACATTTCTCTCCAGGTGTCGAGGAC	862
698	GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet	717
861	GGACGCTTGACGACAGTCACGGCGGTACGGTGACTTCCGCAACACCGGTGGTGATG	802
718	ThrSerAsnLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr	735
801	ACCTCCAACTCGGTTCGGCGCAGATCCAGGAGCTGGCGGC-----GAC	757
736	MetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu	755
756	CGGAGGCGCAACGTGCCGATGATGACCGCGTCAATGCGCACTTCGTCGCCGAATTC	697
756	LeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLys	775
696	ATCAACGGATCCGACGAAGTGGTGTTCGAGCGCTGGCTCGCGACGAGATCCCGGC	637
776	ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu	795
636	ATCCGCGAGATCCAGCTCGTTCGCTCGCGAAGCGCTTCGGCGCGAGCTGAGCCTG	577
796	AlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly	815
576	GAATGACCGAGGAGCGCTGGACAGCTGATTCGCTGGCTTCGACCCCGTCTATGGC	517
816	AlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetVal	835
516	GCACGCCGCTGAAGCGGCCATCCAGCGCTGGATCGAGAACCCGCTGGCGCACTGATC	457
836	ValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeu	855
456	CTGGCCGGCAATTCGGCGCGGTTGCCAGTATCTCGCGAAGGTGGAGGCGGACGAGATC	397
856	ValTyrArgValGluSerGlyGlyLeu-----ValAspAlaSerThrGlyLysLys	872
396	GTCTT-CGCTGACCTCTTCGGGGCGGACAGGGAAGCCCCGCTTCGGCGGG-----	344
873	SerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAla	899
343	-----CTTTTTCATGG-GGCTCCGGTAGGCGGAGAGCGCTGCGCGA	303

APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana

1. AFFRICANTI: isonigalis, john

TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 REFERENCE: 00786/361002

CURRENT APPLICATION NUMBER: US/09/199,637A
 CURRENT FILING DATE: 1998-11-25
 PRIOR APPLICATION NUMBER: 60/066,517

PRIOR FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437

SOFTWARE: FastSeq for Windows Version 4.0
 Q ID NO 1

LENGTH: 42235

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

FEATURE:

NAME/KEY: variation

LOCATION: (1)...(42235)

OTHER INFORMATION: N is any nucleic acid.

09-199-637A-1

Alignment Scores:

i. No.: 1.37e-205 Length: 42235
 Score: 2133.00 Matches: 440
 Percent Similarity: 68.90% Conservative: 178
 Local Similarity: 49.05% Mismatches: 257
 Indels: 24
 Gaps: 8

09-812-350-17 (1-911) x US-09-199-637A-1 (1-42235)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20

26176 ATCGAATAGACCGTTTGGACGACAGCTCAACTCGCGCTCTCCGACGCCAGTCCCTG 26117

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40

26116 GCCGTGGCCATGACCATCCGCCATCGGCGTGCACCTGTTCCGCTTCCGCTGCTCGAG 26057

41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaGln 60

26056 CAGCAAGCGGCTTCGATCAAGCCCTGCTGATCAGCTCGGCTTCGATATCGCGCCCTG 25997

61 SerAlaGluArgValIleAsnGlnAlaLeuLysLeuProSerGlnSerProPro 80

25996 CGCAGCGGC-----CTCAACAAGAACTCGACGCGCTCCGAGATCCAGAGCCGAC 25943

81 AspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAlaGln 100

25942 GCGCAGCTGAACCTGCTCCAGGATCTCGACGCTGCTCAACGAGCTGACCGCTGGCC 25883

101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 119

25882 CAGCAGAGGGCGACCAAGTTCATCTCCAGCAGCTGGTATGTCGCCGCGATGACGAG 25823

120 AspSerGlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139

25822 AACACAGGCTCGCAAGCTGCTCGGCCAGGGGCTGCGCGAAGCGCTGAGAAAT 25763

140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159

25762 GCGGTGGCCAACTGCGTGC-----GGCAAGCGGTGAACGACCCGACCTCGAGAG 25709

160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177

25708 TCGCGCCAGGCGCTGGACAAGTACACCTGCGACATGACCAAGCGCGCGAGGAGCAAG 25649

178 LeuAspProValIleGlyArgAspGluGluIleArgArgValAlaArgIleLeuSerArg 197

25648 CTCGACCCCGGTGATCGTCCGACGACGAGATCCGCGGACCATCCAGGCTCCGAGCGG 25589

198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217

25588 CGGACCAAGAACAAACCCCGTGTGATCGCGGACCCGCGGCTCGGCAAGCCGCCATCGTC 25529

QY 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 DB 25528 GAGGGCTGGCCAGCGCATCATCAACGGCGAGTGCAGGCGCTCAAGGACAGCGC 25469
 QY 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
 DB 25468 CTGCTGGCCCTGGACATGGGGGGCTGATCGCGGTGCCAAGTTCGCGCGAGTTCGAG 25409
 QY 258 GluArgLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 DB 25408 GAACGCTGAGGCGGTCTCTCAACGAACCTGGGCAAGGAGCGCGGTCTCATCTGTT 25349
 QY 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAla 297
 DB 25348 ATCGAGAACTGCACCATGTGCGCGCGCGCAAGGCGGAGGTGCCATGGACCCCGC 25289
 QY 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
 DB 25288 AACATGCTCAAGCGCGCTCTGGCGCGCGCGAGCTGCTGCGTGGTGTCTACTACCTC 25229
 QY 318 GluGlyTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnVal 337
 DB 25228 GACGAGTATCGCCAGTATCATCGAGAAGATGCGCGCTGCGAGCGCGCTTCCAGAAGGT 25169
 QY 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
 DB 25168 CTGGTGGAGCAACCGCGAGGAGACCATTCGCTCTCCGTGGCTCAAGGACGC 25109
 QY 358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeu 377
 DB 25108 TATGAAGTGCACACCGGCTGAGCATCACCGCGCGCGATCATCGCGCGCGCCAGCTG 25049
 QY 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 DB 25048 TCGCACCGCTCATCATCACCGTGCACACTGCGGACACTGCGGACAGGCTATCGACCT 24989
 QY 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 DB 24988 GCCGCCAGCCCATCCGATCGGATCGACTTCAAGCGCGAGGAACTGGATCGTCTCGAC 24929
 QY 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 DB 24928 CGTGGCTGTGATCGAGTGAAGTGCAGCGCGAGCGCTGAAGAGGAGAGCAGCAGAGCC 24869
 QY 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGln 457
 DB 24868 ACCAGGAGCCCTGGCGCAAGCTCGAGGAGATATCGTCAAGCTCGAGCGCGATACGCC 24809
 QY 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 DB 24808 GACCTCGAGGAGATCTGGAGTCCGAGAGCGCGAGGTGCGAGGCTCGCGCGAGATCCAG 24749
 QY 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGlnAlaGluArgArgTyrAspLeuAla 497
 DB 24748 CAGAAGATCGAGCAGGCGCAAGCAGAGTGGAGCGCGCGCGCGCAAGGCGCGCTCGAG 24689
 QY 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 DB 24688 AGCATGGCGCGCTCCAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCA 24629
 QY 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 DB 24628 GACCAGCAGCGCAAGACCGAGACCATGCTGCTGGCGCAACAGGTGACCGCGAGGAATC 24569
 QY 538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 DB 24568 GCGAGTGTTCCTCAAGTGCAGCGGTATCCCGGTGTCCAGAGATGCTCGAGGCGCGAGCC 24509
 QY 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 DB 24508 GAGAAGTGTGCGCATGGAGCGAGGAGTGCATCGCGCGAGTGCATCGCGCGAGGAGCGG 24449
 QY 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuArgAlaGlnGln 597

```

24448 GTAGTCCGCGTGTCCAAAGCGCGTGGCGGCTCGCGCGCGCGCTCGCGCATCCGACCGG 24389
598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyValGlyThrGluLeuAlaLys 617
24388 CCGAGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24329
618 AlaLeuAlaGluGlnLeuPheAspGluAsnLeuValArgLeuValArgLeuValArgLeu 637
24328 GCGCTGCGCGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24269
638 TyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProProGlyTyrValGlyHis 657
24268 TCCATGAGAAACACTCGTGGCGCGCGCTGATCGCGCGCTCGCGCTCGCGCTCGCGCTTC 24209
658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValLeuLeuPhe 677
24208 GAGGAAGCGGCTACCTGACGAGGCGATCCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCT 24149
678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
24148 GACGAGGTGGAGAGGCCATCCGATGATTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24089
698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValLeuLeuMet 717
24088 GACGCGCTGACCGACAGTCAAGGCGGCTGACGCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 24029
718 ThrSerAsnLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr 735
24028 ACCTCAACCTCGTTCGCGCGAGATCCAGGAGCTGGCGGC-----GAC 23984
736 MetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu 755
23983 CCGAGGCGCAACGTCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 23924
756 LeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeuArgLys 775
23923 ATCAACCGATGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 23864
776 ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu 795
23863 ATCCGCGAGATCAGTCTCGCTCGCTCGCAAGCGCTGCGCGAGCGCGAGTGTGAGTGTG 23804
796 AlaValThrAspAlaLeuAspTyrLeuLeuAlaGluSerTyrAspProValTyrGly 815
23803 GAACGAGCGAGGCGCTGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 23744
816 AlaArgProLeuArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetVal 835
23743 GCACGCGCGCTGACGCGCGCTGACGCGCGCTGACGCGCGCTGACGCGCGCTGACGCGCG 23684
836 ValArgGluGluLeuAspGluAsnSerThrValTyrLeuAspAlaGlyAlaGlyAspLeu 855
23683 CTGGCGCGGCAATTCGCGCGCGCTGACGCGCGCTGACGCGCGCTGACGCGCGCTGACG 23624
856 ValTyrArgValGluSerClyGlyLeu-----ValAspAlaSerThrGlyLysLys 872
23623 GTCCTT-CGCTT-CGCTT-CGCTT-CGCTT-CGCTT-CGCTT-CGCTT-CGCTT-CGCTT- 23571
873 SerAspValLeuLeuHisIleAlaAsnGlyProLysArgSerAspAlaAla 889
23570 -----CTTTTCATGG-CGCTTCGCGTAGCGCGGAGCGCTGCGCGCA 23530

```

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4842
; LENGTH: 2580
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4842

Alignment Scores:
Pred. No.: 1,92e-207 Length: 2580
Score: 2132.00 Matches: 427
Percent Similarity: 69.95% Conservative: 176
Best Local Similarity: 49.54% Mismatches: 245
Query Match: 46.86% Indels: 14
DB: 4 Gaps: 6

US-09-812-350-17 (1-911) x US-09-252-991A-4842 (1-2580)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaHisGluLeu 20
Db 16 ATGGAATAGACCGCTTTGACCAAGCTGCACTGCGGCTCTCCGACGCCAGTCCCTG 75
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 76 GCGTTGGCATGACCATCCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCG 135
QY 41 AspProThrGlyLysPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Db 136 CAGCAAGCGGCTTCGATCAAGCCCTGCTGATCAGGTGCGTTCGATATCGCGCGCTG 195
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
Db 196 CGCAGCGCGC-----CTCAACAAGAACTCGACGCGCTGCGCAAGATCCAGAGCCG 249
QY 81 AspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAlaAlaGln 100
Db 250 GCGAGCTGAGTCTGCTCCAGGATTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 309
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu---Glu 119
Db 310 CAGCAGAGGCGGACCAAGTCTATCTCCAGCGAGCTGATGCTGCTGCTGCTGCTGCTGCT 369
QY 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
Db 370 AACACCGAGTCTGCGCAAGTCTGCTGCGCGCGCTGCTGCGCGCGCTGCTGCGCGCG 429
QY 140 GluValGluLysLeuArgGlyLysGlyLysValLysValGluSerAlaSerGlyAspThr 159
Db 430 GCGTGGCGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
QY 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
Db 484 TCAGCGCGAGCGCTGCGCAAGTCTATCCAGCGAGTCTGATGACCAAGCGCGCGCGAGG 543
QY 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
Db 544 CTGATCCGCTGATCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 603
QY 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
Db 604 CGGACCAAGAACACACCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 663
QY 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
Db 664 GAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
QY 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
Db 724 CTGCTGCGCGCTGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783

```

09-252-991A-4842
 Sequence 4842, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A

618	AlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGlu	637
1864	GCCTGCGCCAGTTCCTTCGATACCGAGGAGCGCTGCTGCGGATCGATATGTCGAG	1923
638	TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis	657
1924	TTCATGGAGAAACACTCGGTGGCCGCCCTGATCGGCGCGCTCCGGGCTACGTCGGCTTT	1983
658	GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe	677
1984	GAGGAAGCGGCTACTCTGACCGAGGCGATCGCCGARGCCCTACTCGGTGGGTGCTGCTG	2043
678	AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp	697
2044	GACGAGGTGGAGAAGCCCATCCGGATGTATTCAACATTCTCTCCAGGTGCTCGAGGAC	2103
698	GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet	717
2104	GACGCGCTCAGCGACAGTCACCGGGCGTACGGTGACTTCGCGAACACCGTGGTGGTATG	2163
718	ThrSerAsnLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr	735
2164	ACCTCCAACTCGGTTCGGCACAGATCCAGGAGCTGGCCGGC-----GAC	2208
736	MetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu	755
2209	CGCAGGCGCAACGTGCCGAGTCATGATGACGCGGTCAATGCACACTTCGTCGGAATTC	2268
756	LeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLys	775
2269	ATCAACCGGATCGAGAAATGGTGATATTCAGCGCGTGGCTCGCGACGATCGCCGCG	2328
776	ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu	795
2329	ATCGCGGAGATCCAGCTCGGTGCGCTGGCAAGCGCCTGGCCGAGCGCGAGCTGAGCCTG	2388
796	AlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly	815
2389	GAACTGACCGAGGAGGCGCTGCACAAGCTGATCGCCGTCCGCTTCGACCCCGTCTATGCG	2448
816	AlaArgProIleArgArgTrrpMetGluLysLysValValThrGluLeuSerLysMetVal	835
2449	GCACGCCGCTGAAGCGGGCCATCCAGCGCTGGATCGAGAACCCGCTGGCGCACTGATC	2508
836	ValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeu	855
2509	CTGGCCGCCAAATTCGCGCGGGTCCGAGTATCTCGCGGAAGGTGGAGGACGAGATC	2568
856	ValTyr	857
2569	GTCTTC	2574

RESULT 13
US-09-252-991A-4897/c
; Sequence 4897, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO
/ TITLE OF INVENTION: ASRUGINOSA FOR DIAGNOS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 4897
/ LENGTH: 2970
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4897

```

Alignment Scores:

2.41e-207 Length: 2970
 d. No.: 2132.00 Matches: 427
 Percent Similarity: 69.95% Conservative: 176
 t Local Similarity: 49.54% Mismatches: 245
 Match: 46.86% Indels: 14
 Gaps: 6

09-812-350-17 (1-911) x US-09-252-991A-4897 (1-2970)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 2949 ATGCGAATAGACCGTTTACACGAGCAAGCTGCAACTGGCGCTCTCCGACGCCCACTCCCTG 2890
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 2889 GCGGTGGCCATGACCATCCGCCCATCCAGCGGTGACCTGCTTCCGCGCTGCTCGAG 2830
 41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 2829 CAGCAAGCGCGTTCGATCAAGCCCTGCTGATGAGTTCGGCTTCGATATCGCGCGCTG 2770
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 2769 GCGAGCGGC-----CTCAACAAGAACTCGACGCGCTGCCGAAGATCCAGAGCCCGACC 2716
 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAlaGln 100
 2715 GCGAGCGTGAACCTCTCCAGGATCTCCAGCGCTGCTCAACAGCTGACCGCTCGCC 2656
 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu---Glu 119
 2655 CAGCAGAAGCGGACCACTTCTCCAGCAGTGTGTTATGCTGGCGCGATGAGACGAG 2596
 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 2595 AACACCGCTCGGCAAGCTGCTCGCGCAGCGCGTGTCCGCAAGCGCGCTCGGAAT 2536
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 2535 GCGTGGCCCAACCTCGCTGGC-----GGCGAAGCGGTGAACGACCCGACGTCGAGGAG 2482
 160 AsnPheGlnAlaLeuLysThrTyArgAspLeuValGluGlnAla-----GlyLys 177
 2481 TCGCGCCAGGCGCTGGACAGTACACCGCTGCATGACATGACCAAGCGCGCGAGAGGCAAG 2422
 178 LeuAspProValIleGlyArgAspGluLysLeuArgValValArgIleLeuSerArg 197
 2421 CTCGATCCGGTGTGCTGGCGAGCAAAATCCCGGACCATCCAGGTCCTGACGCG 2362
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 2361 CGGACCAAGAACACCCGGTCTGATCGCGAACCCCGCGCTCGGCAAGACCGCATCGTC 2302
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 2301 GAGGCGCTGGCCCGACGCGATCATCAACGCGGAAAGTGGCGGACCGCTCAAGGACAAAGCG 2242
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
 2241 CTGCTGGCCCTGGACATGGGGCGCTGATCCCGGTGCCAAGTTCGCGGCGGAGTTCGAG 2182
 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 2181 GAACGCTTGAAGCGAGTCTCAACGAACTGGGCAAGCAGGAGGCGCGGTTCATCTCTC 2122
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 2121 ATCGACGAACATGCACACCTGGTTCGGTCCGCGAAGCGGGAAGGTGCGATGGACGCGCG 2062
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317

Db 2061 AACATGCTCAAGCGCGCTCTGGCGCGCGAGCTGCACTGCGTGGTCTACTACCTC 2002
 QY 318 GluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 337
 Db 2001 GACGAGTATCGCGAGTACATCGAGAGGATCGCGCTGGAGCGCGCTTCAGAGGTG 1942
 QY 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 Db 1941 CTGGTGGACGAAACCGAGCGAGGAGACACGATCGCATCTCGTGGTCTCAAGGAACGC 1882
 QY 358 TyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 Db 1881 TATGAAGTCAACACCGGTGAGCATCACCGCGCGGATCATCGCGCGCCCAAGTGT 1822
 QY 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 Db 1821 TCGACCGCTACATCACCGATCGGCAACTGCGGACAGGCCATCGACTGATCGACGAG 1762
 QY 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 Db 1761 GCGCGCAGCGCATCGCATGAGATCGACTCCAAGCCAGAGAACTGGATCGTCTCGAC 1702
 QY 418 ArgLysArgMetGlnLeuLysGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 Db 1701 CGTCCGCTGATCCAGTGAAGATCGAGCGCGAGCGCTGAAGAGGAAGACGACGAAGCC 1642
 QY 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGln 457
 Db 1641 ACCAGGAAGCGCTGGCCTGAGGAGGATATCGTCAAGCTCGAGCGCGAATACGCC 1582
 QY 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgGluLys 477
 Db 1581 GACCTCGAGGATATCTGGAAGTCCGAGAGGCGGAGTGTGCGCGCAGATCCAG 1522
 QY 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
 Db 1521 CAGAGATCGAGCGCAGCAGGAGATGAGGCGCGCGCGCGCAGGCGGACCTCGAG 1462
 QY 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 Db 1461 AGCATGGCGCATCCAGTACCAGACCATCCCGGACCTGGAAACGCGAGCTGCAGATGTC 1402
 QY 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 Db 1401 GACCAGCAGCGCAAGACCGAGAACCATGTTGCTGTCGCAACAGGTGACCGACGAGAAATC 1342
 QY 538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 Db 1341 GCGAAGTGGTTCAGAGTGGACCGGTATCCCGGTGCAAGATGCTCGAGGCGGAGCGC 1282
 QY 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 Db 1281 GAGAGCTGCTGGCATGAGCAGCAGGAGCTGATCGCGAGTGTGATCGCGCAGGAGCGG 1222
 QY 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 Db 1221 GTAGTCCCGTGTCAACGCGCTGCGCGCTGCGCGCGCGCTCGCGATCCGAAACCGG 1162
 QY 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
 Db 1161 CCGAGCGGTCTGTTCTCTCTCTCTCGCGCGAGCGGCTGGTGGGACGAGCTGTGTCAG 1102
 QY 618 AlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGlu 637
 Db 1101 GCGCTGGCGAGTCTCTCTCTATACCGAGGAGGCGCTGGTGGGATCGATATGTCTCCAG 1042
 QY 638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHis 657
 Db 1041 TTCATGGAGAAACATCTGGTGGCGCGCTGATCGCGCGCGCTCGCGGTCTCGCGCTTT 982
 QY 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 Db 981 GAGGAGGCGCGTACTGACCGAGCGGATCCGCCCAAGCCCTACTCGGTGTGTGTGTCG 922

86 SerSerSerLeuIleLysValIleArgArgAlaGlnAlaalaadGlnLysSerArgGlyAsp 101
Db :
262 AGTGTCAAAA CAAATGAATTATTAGACA CAAGCCGACAGAAATATATGAAGCTCTTAAGAAGT 321
QY :
106 ThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluaspSerGlnIleArgasp 125
Db :
332 GAGTTTATTTCTATGGAGCATATATTACTGTCGACCATTAGATACTGATGAAATCTACTCAA 381
QY :
126 LeuLeuAanGluValGlyValAlaIleThrAlaArgValLysSerGluValGluLysLeuArg 145
Db :
382 -----AAATGGGTCCGC AATAAGTAGAGTGATATAAAGAAATTAATACTAAAGTTAGA 435
QY :
146 GlyLysGluGlyLysLysValGluSerAlaSerGlyAspThrAanPheGlnAlaLeuLys 165
Db :
436 GGG-----GGAAATCATGTGCATCTCAAAATCCAGAAAGTAAACTATCGAAGCTTTAGAA 489
QY :
166 ThrTyrgLyArgAspLeuValGluGlnAla-----GlyLysLeuaspProValledgly 183
Db :
490 AAATATGTGTGTGATCTTGTAGAGAAAGTAGTCGCAAGGTAAAAATGACCCCTGTCACTCGT 549
QY :
184 ArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThryrLysAanAenPro 203
Db :
550 AGAGATCGAGAAATACGTATACATTCGTATATTAAGTTCGAAAAACAAAAATTAATCCG 609
QY :
204 ValIleuIleGlyGluProGlyValGlyLysThrAlaValVaIGluGlyLeuAlaGlnArg 223
Db :
610 GTTTTAAATCGGTGAACAGGTGTAGGTAAAACTGCAATTTGTTGAAGTTTTAGCAACAAGA 669
QY :
224 IleValLysGlyAspValProAanSerLeuThrAspValArgLeuLysSerLeuaspMet 243
Db :
670 ATTGTAGCTAAAGACGTTCCAGAAATCATCTACTCGATAAAACGATTTTCGAGCTTGATTTA 729
QY :
244 GlyAlaLeuValAlaGlyAlaLysTyraArgGlyGluPheGluGluArgLeuLysSerVal 263
Db :
730 AGTCCCTAGTTGCAGTGCARAATATAGAGTGGAATTTGAAGAACGTTTAAAGCAGCTA 789
QY :
264 LeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGluIleHisLeu 283
Db :
790 TTGAAAGAGTAAAGAAATCTGAAGGTCCGAATCATCTTATTCTATGATGAAATTCATATG 849
QY :
284 ValLeuGlyAlaglyLysThrGluGlySerMetAspAlaAlaAanLeuPheLysProMet 303
Db :
850 TTAGTAGGCCGTGGTAAAACTGATGTGTCTATGACGCGCAGGAACCATCTTGAAGCAATG 909
QY :
304 LeuAlaArgGlyGlnLeuArgCysIleGlyAlaIleThrLeuGluGluTyraArgLysTyrr 323
Db :
910 CTTCGGCGTGGTGAATTCGATGTATAGTGTCTACGACATTAATGAATATCTCGTAATAT 969
QY :
324 ValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyraAlaGluProSer 343
Db :
970 ATTCAAAAGAAATTCAGCAATTAGAACGTGCTTTTCCAAAAGTTGGGGTCAGTGAACCTGAT 1029
QY :
344 ValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrrGluGlyHisGly 363
Db :
1030 GTTCAGAAATCAATATCAATTTTACGTGTCTTAAAGAACGTTATGAGGTCTACCATGGT 1089
QY :
364 ValArgIleGlnAspArgAlaLeuIleAenAlaAlaGlnLeuSerAlaAatqTyrrIleThr 383
Db :
1090 GTAGTATACAGACCGTGCATCTTTGTCGCGGTGAGCTATCTGTATAGATATATTATACC 1149
QY :
384 GlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAlaAanValArg 403
Db :
1150 GATAGATTTTTCAGATAAAGCTATAGATTTAGTTCGACCAAGCTTTGTGCGACAATACGT 1209
QY :
404 ValGlnLeuAspSerGlnProGluGluIleAspAanLeuGluArgLysArgMetGlnLeu 423
Db :
1210 ACAGAAATGGCATCAATCCAAGAGTTAGCACCAAGTAAATCTCGTAGTAATGCAATTA 1269
QY :
424 GluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAlaAargLeuIle 443
Db :
1270 GAATTTGAGAAAGCCCTTTGAANAATGAATCTGATAACGCAAGTAAACATCTCGTAGTAATGCAATTA 1329
QY :
444 GluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrr 463

1330 GAGCTTCAAGAGGAAATATCTAATGAAAAAGAAAAACAATCCTCTTTTAAATCAAGTGT 1389
464 ArgLysGluLysGluArgLysGluArgLysGluArgLysGluArgLysGluArgLysGlu 483
1390 GAACAAGAAAAAGAAAAA-----ATTGCTAAAGTACAAAGAAAAAAGCTGCTGAAT 1440
484 MetPheSer-----LeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
1441 GATAGTAGTTCGCTCAAGCTCTTGAAGATGCACAAACAGAGGTAATCTTCAAAAAGCAGCA 1500
501 AspLeuArgTyrGlyAlaLeuGlnGluValGluSerAlaLeuAlaGlnLeuGlu----- 518
1501 GAATTACAATATGTTCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCA 1560
519 -----GlyThrSerSerGluGluGluValMetLeuThrGluAsnValGlyProGlu 535
1561 TTTCAGATGAACACTGGCAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1620
536 HisLeuAlaGluValValSerArgTyrThrGlyLeuProValThrArgLeuGlyGlnAsn 555
1621 GAATAGGCGACATAGTATGATCAATGCAGCAGTATACCTGTATCTAAACTTGTGGAACA 1680
556 GluLysGluArgLeuLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsn 575
1681 GAAAGAGAAAGCTATTAAAGTTTAAAGTTCATCTTACACAAACGCTAGTAGGTCAAGAT 1740
576 GlnAlaValAsnAlaValSerGluAlaLeuLeuArgSerArgAlaGlyLeuGlyArgAla 595
1741 AAAGCAGTTGATTAGTATGATCAGCAGCAGTATGAGCAGCAGTATGAGCAGTATGAGC 1800
596 GlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeu 615
1801 AATAGACCAATCGGAAGTTCTTATTTCTTAGCCTTACTGAGTAGGTAAACTGAATTA 1860
616 AlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgLysLeuMet 635
1861 GCAAAATCGCTTCATCCTTCATCCTTCATCCTTCATCCTTCATCCTTCATCCTTCATC 1920
636 SerGluTyrMetGluGlnHisSerValSerArgLeuLeuGlyValArgProGlyTyrVal 655
1921 AGCGATATATGAAAAACATGCTGTATCAGCTTAAATGTTGTCACCTCCAGCTTATGTA 1980
656 GlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValle 675
1981 GGTCCACATGAAGTGTCAATTAACCTGAAGCAGTTAGACGTAATCCATCTCAGTTAT 2040
676 LeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuValLeu 695
2041 TTGTTAGACGAAGTTGAAAAAGCAGCAGTATGTTTAAATGTTATCTTCAAAATCTA 2100
696 AspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValle 715
2101 GATGAAGTTCGCTTACGATCTTAAAGTAGAGTGTGACTTTAAATAACCATATATC 2160
716 IleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeu-----ThrGlyLys 733
2161 ATCATGACGTAGTAATATGTTGTTCAACAGTATTACTTGAAGATGTAAGATGCTGCTGA 2220
734 ValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgPro 753
2221 ATTAGTGATGATACAGAGAAA---GCAGTTATGGACAGTCTACATGCTACATCTCAAACT 2277
754 GluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeu 773
2278 GAAATATTAATCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2337
774 ArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgLysVal 793
2338 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2397
794 AlaLeuAlaValThrAspAlaLeuAspTyrLeuLeuAlaGluSerTyrAspProVal 813

Db 2398 TCAATTGAAGTGACAGAGAGCGGAAAAAATGGCTAGGTGAAGAGCGTATGACACCAA 2457
Qy 814 TyrGlyAlaArgProfileArgArgTyrMetGluLysLysValValThrGluLeuSerLys 833
Db 2458 TTTGCTGACAGACCATTAACACGCTTTTGTTCACGACAAATAGAAACTCCAATTGCACGT 2517
Qy 834 MetValValArgGluGluLeuLeuAspGluAsnSerThrValTyrIleAsp 849
Db 2518 ATGATGATTAAAGAAAGTCTACCTGAAGGTACATAATAATTAAAGTAGAT 2565
RESULT 15
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t

NAME/KEY: misc feature
LOCATION: (225001)..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (240001)..(250000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (255001)..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (270001)..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (285001)..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (300001)..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (315001)..(330000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (330001)..(345000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature

LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature

Alignment Scores:

Pred. No.:	1,11e-180	Length:	1230025
Score:	1912.00	Matches:	404
Percent Similarity:	66.24%	Conservative:	171
Best Local Similarity:	46.54%	Mismatches:	271
Query Match:	42.02%	Indels:	22

09-812-350-17 (1-911) x US-09-198-452A-1 (1-1230025)

4 Gaps: 8

4 GluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAlaValAsn 23
 188339 GAGAAATTTTCGATGCTCTCTGAGCTTTAGAGAGGCTTTTCGAACTTCTGCTAAATCT 188398
 24 AlaGlyHisAlaGlnPheProLeuHisLeuAlaGlyAlaLeuIleSerAspProThr 43
 188399 TCGAACATACCTATGTCACAGAAATACCTATTACTGGCTTTATTAGAAATACAGAG 188458
 44 GlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAlaAlaGlnSerAlaGlu 63
 188459 TCTCTTTTATTGTAATTAAGACATTCATCGGAACTCTGCTGCTTTTGGAGT 188503
 64 ArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro----- 79
 188504 ---TGTCTCAATACGCGATTAAGATGCGCTCTCACGAGAGCCGACTGTAGTTGAAGGA 188560
 80 ProAspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAla 99
 188561 GAGGTGATCTTAACCTCTCGGGTTTACAAACCTCTTAGGGATGCCAAACAGAG 188620
 100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
 188621 GCAAGACATTAGGAGATGAATACATTTCTGGAGATCATCTGCTGCTTTTGGAGT 188680
 120 AspSerGln-----IleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
 188681 TCAACAAAGAGCCCTTTAATTTCTTGAAGCAACAAAGTTAGTTTAAAGATCTT 188740
 138 LysSerGluValGluLysLysArgGlyLysGluLysValLysValGluSerAlaSerGly 157
 188741 AAGAATCTGATTACTAAATACGA-----CGAGGAATCGTATGATGCGCAAGCGCT 188794
 158 AspThrAsnPheGlnAlaLeuLysThrTyArgAspLeuValGluAla----- 175
 188795 GAAAGTAATTTTCAGGGTTTGAAGATTAATCTAATAATTTTAAACAGCATTAGCTGTA 188854
 176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgValValArgIleLeu 195
 188855 GGTAACTGGATCTCTGATCGTAGAGATGAAGAAATTCGTAGAACCATCCAGTGTCT 188914
 196 SerArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
 188915 TCCCGTAGAACTAAATAAACCCTATGCTATTGCTGAGCGGGGTAGGAAATCTGCT 188974
 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
 188975 ATAGCAGAGGATTAAGCTTTAGGCTTTATCCAGGGTGTATCTCTGAATCTCTCAAAGGT 189034
 236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyValLysThrArgGlyGlu 255
 189035 AAACAGCTTTATGCTTAGATATGGAGCTTTGATTGCGAGGAGCTAAGTATCGAGGTAG 189094
 256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIle 275
 189095 TTTGAAGAAAGACTAAGAGTGTTTTAAAGATGTAGATCTCGAGATGGCGAGCACATT 189154
 276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
 189155 ATCTTTATTGATGAGTGCTACTCTTTGTTGAGCAGGAGCTACTGTAGTGGAGCTTGGAT 189214
 296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
 189215 GCTGCAATCTTTAAAGCTTCATTAGCAAGAGGAGCGCTACACTGTATTGGCGGAGC 189274
 316 ThrLeuGluGluTyArgLysThrValGluLysAspAlaAlaPheGluArgArgPheGln 335
 189275 ACTTTGATGATGATCAGAGATATTAAGAAAGATGCTGCTTTTGAACGCTGATTTTCCAG 189334
 336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355

189335 CCTATTTTGTGACAGAGCTCTCTTTGGAGATGCTCTTTTATTTCTCTGCGACTAAGA 189394
 356 GluLysTyThrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
 189395 GAAATAATATGAAATTTTCCATGGAGTCAAGATTAAGAGGGGCTTTGAATGCGCGATC 189454
 376 GlnLeuSerAlaArgTyThrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
 189455 CTACTTTCTCTATCGTTATATCCAGATCGCTTTCTTCCAGTAAGCTATCGATTGATA 189514
 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
 189515 GATCAAGCGCAAGTTTAATTCGATCAAAATGGTAGTCTCTCTCTCTCTATTTATGAA 189574
 416 LeuGluArgLysArgMetGlnLeuIleGluLeuHisAlaLeuGluArgGluLysAsp 435
 189575 AAGGAGAGAGCTGCTGCTGTTGATGCTTGAACAGAGCTATAAACGCGAGCAATCT 189634
 436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLys 455
 189635 CTTTCTCTATCAAGAGAGCGGATGCTATGACAGATCTATAGATGCTTTGACAGAGGAA 189694
 456 LeuGlnProLeuThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArg 475
 189695 TTAGCATCTCTACGTTTGGGTGGGATGAAGAGAAGTTGATTTCCGGGCTCAAGGAA 189754
 476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyAsp 495
 189755 AAAGAGATTCCTTGGAAAGTATGAATTTCTTGAAGAGAGCGGAGCGTGTTCAGAC 189814
 496 LeuAlaArgAlaAlaAspLeuArgTyGlyAlaIleGlnGluValGluSerAlaIleAla 515
 189815 TATAATCGTGTAGCTGAGCTTCGTTATAGTTTAAATTTCCCAACTTGAAGAAGAAATCAAA 189874
 516 GlnLeuGluGlyThr---SerSerGluGluAsnValMetLeuThrGluAsnValGlyPro 534
 189875 CAGATGAAGCCCTTTTAAATCAAGAGATAACCGTCTCTCTCAAGAAGAAAGTTGACGAG 189934
 535 GluHisIleAlaGluValValSerArgTyThrGlyIleProValThrArgLeuGlyGln 554
 189935 CGATTGATTGCGCAAGTGTAGTAAATGGAAGAGGATTCCTGTGCAAAAAGCTAGAA 189994
 555 AsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGln 574
 189995 GCGAGAGCTGAGAACTGTTAAATCTTGAAGATCTTGAAGAAGCTGTGTAGGACAG 190054
 575 AsnGluAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArg 594
 190055 CTTTTCAGCTCTCTCGGTAGTGTATCTTATTCGTGCTGCACGTTAGGTTTAAATGAT 190114
 595 AlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlu 614
 190115 CCTCAAGCTCTCTAGAGCTCTTTTATTTTAGGCCCAACAGGGGTAGAAAACCGAG 190174
 615 LeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAsp 634
 190175 CTTGCAAAAGCTCTTCAGATCTCTTTTCAATAAAGAGAAAGCTATGTCGCTTCGAT 190234
 635 MetSerGluTyMetGluGlnHisSerValSerArgLeuIleGlyValaProProGlyTy 654
 190235 ATGTAGATATATGAAGAAGCATTCATTTCCAGCTTATAGGATCTTCTCCAGGGTAT 190294
 655 ValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgProTyCysVal 674
 190295 GTGGTTATGAGGAAGTGGAGTCTTCTGAGGCTCTTCGACGAGCTCCCTATTTCAGTA 190354
 675 IleLeuPheAspGluValGluLysAlaHisAlaValaPheAsnThrLeuLeuGlnVal 694
 190355 GTTCTCTTTGATGAGTGAAGAAAGCAGATAAGGAAGTTCTAAATATCTCTTTTACAGGTT 190414
 695 LeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValaAspPheArgAsnSerVal 714

```

arch completed: February 13, 2004, 08:40:35
time : 13332 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus p2n model

on: February 13, 2004, 04:00:27 ; Search time 713 Seconds
(without alignments)
4706.577 Million cell updates/sec

le: US-09-812-350-17
fect score: 4550
uence: 1 MNPEKFTKNTIATIAHEL.....VKMKRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

arched: 2449703 seqs, 1841816367 residues

al number of hits satisfying chosen parameters: 4899406

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

mand line parameters:

DBL=frame+p2n.model -DEV=xlp
/cgcn2_1/USPTO.spool/p/US09812350/runat_10022004_101517_13444/app_query.fasta_1.1095
Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
DCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
KLEN=2000000000 -USAP=US09812350 @CGN 1 1.85 @runat 10022004_101517_13444
PU=6 -ICPU=3 -NO WARN -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
NGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
APOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : Published Applications NA:
1: /cgcn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgcn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgcn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgcn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgcn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgcn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgcn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgcn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgcn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgcn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgcn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgcn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgcn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:
14: /cgcn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgcn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgcn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cgcn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cgcn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	ID	Description
-------	-------	--------	----	-------------

1	4545	99.9	2736	10	US-09-938-842A-235	Sequence 235, App
2	4545	99.9	2736	12	US-09-938-842A-235	Sequence 235, App
3	3657	80.4	3942	12	US-10-310-154-113	Sequence 113, App
4	2315	50.9	2643	12	US-10-369-493-43734	Sequence 43734, A
5	2288	50.3	2619	14	US-10-369-493-26449	Sequence 26449, A
6	2288	50.3	2619	14	US-10-047-260-37	Sequence 37, Appl
7	2288	50.3	3180	12	US-10-310-154-116	Sequence 116, App
8	2282	50.2	2637	15	US-10-156-761-7220	Sequence 7220, Ap
9	2282	50.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
10	2264.5	49.8	2571	12	US-10-369-493-44175	Sequence 44175, A
11	2247	49.4	2586	12	US-10-369-493-43449	Sequence 43449, A
12	2245.5	49.4	43980	12	US-10-398-221-5	Sequence 5, Appli
13	2245.5	49.4	4011208	12	US-10-398-221-2058	Sequence 2058, Ap
14	2234.5	49.1	2715	12	US-10-369-493-46242	Sequence 46242, A
15	2228	48.9	2588	12	US-10-369-493-44640	Sequence 44640, A
16	2225	48.9	2587	12	US-10-369-493-35669	Sequence 35669, A
17	2224	48.9	2580	12	US-10-369-493-40521	Sequence 40521, A
18	2214.5	48.7	2721	12	US-10-369-493-27677	Sequence 27677, A
19	2204	48.4	2586	9	US-09-815-242-9990	Sequence 9990, Ap
20	2199.5	48.3	2574	12	US-10-369-493-34295	Sequence 34295, A
21	2198	48.3	2577	12	US-10-369-493-35080	Sequence 35080, A
22	2196.5	48.3	2529	12	US-10-369-493-31518	Sequence 31518, A
23	2194	48.2	2574	13	US-10-301-997-46	Sequence 46, Appl
24	2194	48.2	4248	12	US-10-310-154-115	Sequence 115, App
25	2193	48.2	2574	9	US-09-815-242-6161	Sequence 6161, Ap
26	2193	48.2	2574	10	US-09-960-428-20	Sequence 20, Appl
27	2193	48.2	2574	12	US-10-369-493-47195	Sequence 47195, A
28	2184.5	48.0	2574	12	US-10-369-493-41513	Sequence 41513, A
29	2182.5	48.0	2697	12	US-10-369-493-26456	Sequence 26456, A
30	2174.5	47.8	2574	12	US-10-369-493-39286	Sequence 39286, A
31	2166.5	47.6	2547	12	US-10-369-493-39663	Sequence 39663, A
32	2166.5	47.6	2547	12	US-10-369-493-40020	Sequence 40020, A
33	2162.5	47.5	2571	9	US-09-815-242-7037	Sequence 7037, Ap
34	2162.5	47.5	1830121	12	US-10-329-670-1	Sequence 1, Appli
35	2162.5	47.5	1830121	15	US-10-329-670-1	Sequence 1, Appli
36	2158.5	47.4	2586	12	US-10-369-493-41201	Sequence 41201, A
37	2158.5	47.4	3000	12	US-10-310-154-117	Sequence 117, App
38	2151.5	47.3	2556	12	US-10-369-493-31783	Sequence 31783, A
39	2150.5	47.3	3309400	10	US-09-738-626-1	Sequence 1, Appli
40	2148.5	47.2	2571	12	US-10-369-493-32853	Sequence 32853, A
41	2146.5	47.2	2571	12	US-10-369-493-33061	Sequence 33061, A
42	2145	47.1	2556	10	US-09-738-626-3499	Sequence 3499, Ap
43	2141.5	47.1	2607	9	US-09-815-242-4175	Sequence 4175, Ap
44	2141.5	47.1	2649	9	US-09-815-242-8428	Sequence 8428, Ap
45	2141	47.1	2436	12	US-10-369-493-43090	Sequence 43090, A

ALIGNMENTS

RESULT 1

US-09-938-842A-235
; Sequence 235, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepe, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT FILING DATE: 2001-08-24
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 235
; LENGTH: 2736
; TYPE: DNA

ORGANISM: Arabidopsis thaliana
09-938-842A-235

Alignment Scores:

Seq. No.: 0 Length: 2736
Matches: 910
Conservative: 99.89%
Mismatch: 0
Indels: 0
Gaps: 0

09-812-350-17 (1-911) x US-09-938-842A-235 (1-2736)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATCAATCCAGAGAAATTCACACACAGACAAACGACCAATTCCTACAGCTCATGAGCTA 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
61 GCTGTGAATGCAGACATGCTCAATTCCTCCTTTGCAATTTAGCTGCTTTGATCTCT 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaGln 60
121 GATCCACCCGCTAATTTCTCAAGCAATCTCTAGTCCCGGTGGCGAGAACGAGCTCAA 180
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
181 TCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAGCTTCTTCAATCTCTCCACCT 240
81 AspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAlaGln 100
241 GATGATATTCACGAGTCTTAGTCTTATTAAGTCAATTCGTCGCTCAAGCTGCTCAG 300
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
301 AGTCACGAGTGATCTCATTTGGCTGTGTGACCAGTTGATTTAGTGTCTTTTGAAGAT 360
121 SerGlnIleArgAspLeuAsnGlnValGlyValAlaThrAlaArgValLysSerGlu 140
361 TCTCAATCAGGATTTGTGAAGAGTTCGCTGTAGCGACGCGGAGGTAAGTCTGAG 420
141 ValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSerGlyAspThrAsn 160
421 GTTGAGAGCTTCGTGGGAAAGAGGAAAGTGTGAGAGTTCGCTTCAAGGAGACAAAT 480
161 PheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
481 TTTCAGCTTTAAAGACTTATGGAAGAGATTTGGTTGACCAAGCGAGGAGCTTGATCCT 540
181 ValIleGlyArgAspGluGluIleArgArgValValArgileLeuSerArgArgThrLys 200
541 GTGATTTGGTCGTATGAGGAGATTTAGAAGAGTTCGTGAGGATTTCTTTCAGGAGAGACGAAG 600
201 AsnAsnProValIleGluProGlyValGlyLysThrAlaValGluGlyLeu 220
601 AACATCTCTGCTTATTCGAGAGCCAGAGTGTGTTAAACAGCTGTGCTTGAAGGTTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGATTTGGAAGAGATGTGCCCAACAGCTCTTACTGATGTAGATTAATTCG 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGlyGluPheGluGluArgLeu 260
721 TTGGACATGGTGGCTTAGTGTGTGTGCTAAATACCGAGGAGAGGTTTGAAGAAAGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluLysValIleLeuPheIleAspGlu 280
781 AAATCTGTTTGAAGAAAGTTCAGGACGCTGAAGGCAAGTGAATCTCTTTATTGATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTTGGTCTTGGCTGGCAAACTGAAGGTCATGAGTGCAGCTTAATCTGTTTC 900

301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
901 AAGCCATTTAGCTAGAGGCGACCTTCGATGCTTGGTCTACACGCTTGAAGATAC 960
321 ArgLysTyrglyValGluLysAspAlaPheGluArgArgPheGlnGlnValTyrValAla 340
961 AGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGTCTATGTTGCG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLysGluLysTyrglyGly 360
1021 GAGCCAAAGTGTGCTGACACCATTTAGTATCTTTAGAGGACTCAGAGGAAAGTATGAGGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGCTGTGCGAATCCAAAGACAGAGCTCTTATAATGCTGCTCAGCTGTCTGCTCGT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATAACTGCTGCGCATTTACCGGATAAAGCAATTCATTTGGTTGATGAGCTTGTGCG 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspLeuGluArgLysArg 420
1201 AATGTGAGGTCCAGCTTGATGATCAACCTGAAGAGATTCATTAACCTTGAAGGAAAGG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
1261 ATGCAGCTGGAATTTGAATTCAGCTTTCGAAAGGAGAGGATTAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGGTGCGAAAGAGCTTGTATGATGATGATGATGATGATGATGATGATGAT 1380
461 MetLysTyrglyLysGluLysGluValGluIleAspGluIleArgGluLysGlnLysArg 480
1381 ATGAATATCAAGAAAGGAGAAAGAGAGATTTGATGAGATTCGAAGGCTTTAAACAGAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrglyAspLeuAlaArgAlaAla 500
1441 GAAGAGCTCATGTTTCTTTCAGGAGGAGAGAGAGATATGATGATGATGATGATGATGATGAT 1500
501 AspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluTyr 520
1501 GATCTAAGATATGCGCAATTCAGAAAGTGAATCTGCAATTTGCCCAACTTGAAGGAAT 1560
521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
541 ValSerArgTyrglyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
1621 GTGAGCGTTTGGACAGGATTCAGTGTGAGAGCTTGGCCAAATCGAAGGAGAGGTTG 1680
561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAlaAla 580
1681 ATTGGTCTTCTGATAGTGTGATAGCGGTTGTGGGAGAGTTCGAGATCAAGCGGTAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
1741 GTTCTGAGGCAATTTCTAAGTCAAGGAGAGGCTTGGAGGAGGAGGAGGAGGAGGAGGAG 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATTTCTTATTCCTTGGACCAACTGCTGTTGGCAAACTGAGCTGCCCAAGGCTCTTGTCT 1860
621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGlu 640
1861 GAGCAGCTGTTTGTATGATGATAACCTCTTAGTTCGATGATGATGATGATGATGATGATGAT 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGly 660
1921 CAACACTCTGCTCTCTGCTCTCATTTGGGGGACACACAGGATGATGATGATGATGATGATGAT 1980
661 GlyGlnLeuThrGluAlaValArgArgProTyrglyValIleLeuPheAspGluVal 680


```

1981 GGACAACTAACTGAGGCTGTGAGGAGCGACCTTATTGTGTCACTACTTTTGTGATGAAGTG 2040
681 GllulysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeu 700
2041 GAGAAAGGCTCAATGTGCTGCTTCAACACTCTGCTCAAGTTTTGGATGATGTCGATG 2100
701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
2101 ACAGACGGCGAAGCAGGACGACGATTCAGTTCAGGAACTCGGTGATATCATGACATCAAC 2160
721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
2161 CTTGGTGTGAACACCTCTCTGCGGGCTAACTCGGAAAGTAACAATGGAAGTGGCCCGG 2220
741 AspCysValMetArgGluValArgGlyHisPheArgProGluLeuLeuAsnArgLeuAsp 760
2221 GACTGTGTGATGCGGAGGAGTGAAGAAACACTTCACAGCCAGCTCTTTGAACAGCGCTTGAC 2280
761 GllulleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
2281 GAGATTGTGTGTTGCGACCCCTTTCACATGACCAGTTGAGGAAAGTAGCTCGGCTCAA 2340
781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
2341 ATGAAGACGCTGTGCTGCGGCTTGTCTGAAAGAGGAGTGTCTTTGGCAGTCACTGATGCT 2400
801 AlaLeuAspThrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
2401 GCTTTGACATAATCTTTGGCAGAGATTATGACCCGCTGTATGCTGTGCGCTTATAAGG 2460
821 ArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
2461 AGATGGATGAGAGAAAGGTGTCACAGACTGTCAAGATGTTGTGCGCTGAGGAAATC 2520
841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
2521 GATGAAACTCCACTGTTTACATAGATGCGGCGCTGCTGATCTTTGTGTACCGGCTAGAA 2580
861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
2581 AGTGGAGTCTAGTGGAGCTTCAACAGGCAAGAAAGTCAGATGCTGTGATTCATATGCT 2640
881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
2641 AACGGCCAAAGAAAGTGTGACGCTCAGCGGTGAAGAGATGAGATCGAGGAATA 2700
901 GluAspAspAsnGluGluMetIleGluAsp 911
2701 GAAGATCAGCATATGAGGAAATGATCGAGAT 2733

```

JUL 2

Sequence 235, Application US/09938842A
 Publication No. US20040009476A9
 GENERAL INFORMATION:

APPLICANT: Harper, Jeff
 APPLICANT: Kreps, Joel
 APPLICANT: Wang, Xun
 APPLICANT: Zhu, Feng

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 FILE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: SCRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 235

; LENGTH: 2736
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-938-842A-235

Alignment Scores:

Pred. No.: 0 Length: 2736
 Score: 4545.00 Matches: 910
 Percent Similarity: 92.88% Conservatives: 0
 Best Local Similarity: 99.89% Mismatches: 1
 Query Match: 99.89% Indels: 0
 DB: 12 Gaps: 0

US-09-812-350-17 (1-911) x US-09-938-842A-235 (1-2736)

```

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 1 ATGAATCCAGAGAAATTCACACAAAGACAAACGAGACAAATTTGCTACAGCTCATGAGCTA 60
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
DB 61 GCTGTGAATGCGAGGACATGCTCAATTCACCTCTTTGCAATTTAGCTGGTGTCTTGAATCT 120
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
DB 121 GATCCACCGGTATATTTCTTCAAGCAATCTCTAGTGGCGGTGGCGAGAACGAGCTCAA 180
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
DB 181 TCTGCTGAAGAGTATCAATCAAGCTTTGAAGAGCTTCTCTTCAATCTCTCCACT 240
QY 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
DB 241 GATGATATTCACGAGATTCAGTCTTATTAAAGTCATTCGCTGCTGCTCAAGCTGCTCAG 300
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
DB 301 AAGTCACGAGGTGATCTCAATTTGGCTGTTGACCAAGTGTATATGGGTCTTCTTCAAGAT 360
QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
DB 361 TCTCAAAATCAGGAGTATTGTTGAACCAAGTCGCTGAGCGAGCGGAGGAGTAAAGTCTGAG 420
QY 141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
DB 421 GTTGAAGACTTCGTGGGAGAGAGAGGAGGAGTTGTAAGAGTGTCTTCAAGGAGACCAAT 480
QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
DB 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAGGAGCTTATCTCT 540
QY 181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
DB 541 GTGATTTGCTGCTGATGAGGAGATTAGAAGAGTCGCTGAGGATTTCTTTCGAGGAGAACGAAG 600
QY 201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
DB 601 AACATCTCTGCTCTATTGGAGAGCCAGAGTTGTGTAACAGCTGTGTTGAAGGTTTGAAGGTTT 660
QY 221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
DB 661 GCACAAAGGATTTGTAAGAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTG 720
QY 241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluArgLeu 260
DB 721 TTGGACATGGGTGCTTATTGCTGCTTAAATCCGAGGAGAGTTTGAAGAAAGGTTG 780
QY 261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
DB 781 AAATCTGTTTGAAGAAAGTTGAGGACCTGAGGCAAGTGATCTCTTTATTGATGAG 840
QY 281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300

```

841 ATTCAATTGGTCTTGGTGGTGGCAAACTGAGGGTGCATGATGACGCTAATCTGTTTC 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCysValIleGlyValAlaThrThrLeuGluGluTyr 320
901 AAGCCCATGTTACCTAGAGGGCAGCTTCGATGCAATGGTGTCTACAAAGCTTGAAGATAC 960
321 ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrValAla 340
961 AGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAAGCTTATGTTGCG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
1021 GAGCCAAAGTGTGCTGACCACTTAGTATCTTACAGGACTCAAGAGAAAGTATGAGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGGTGTGGAATCCAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTGT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATAAATCTGGTGGCAATTTACCGGATAAGCAATTTGATTTGATGAGGCTTTGTCG 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
1201 AATGTGAGAGTCCAGCTTGATAGTCACTGAGAGATTTGATTAACCTTTGAAAGAGAGG 1260
421 MetGlnLeuGluIleGlnLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
1261 ATGCAGCTCGAAATTTGAATCTCACGCTTTGGAAGAGGAGGATAAAGCCACCAAGCT 1320
441 ArgLeuIleGluValArgLysGlnLeuAspAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGTGGGAAAGAGCTTTGATGACCTGAGAGACAAGCTTCAGCCTCTCAG 1380
461 MetLysTyrArgLysGlnLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
1381 ATGAAATACAGAAAGAGAGAAAGAGAGAAATTTGATGAGATTCGAAAGCTTAAACAGAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
1441 GAAGAGCTCATGTTTTCTTTGAGAGGAGGAGAGCAAGCAAGATATGACCTTGCAGAGCTGCT 1500
501 AspLeuArgTyrGlyValAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1501 GATCTTAGATATGGCGCAATTCAGAGAGTGGAACTTGCATTTGCCCACTTGAAGAACT 1560
521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAAGAGAAATGCTGATGCTCACAGAAACGTTGGSCCTGAACACATTTGCTGAGGTT 1620
541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
1621 GTGAGCGGTGGACAGGGGATTCAGTGACGAGACTTGGCCAAATGAGAGGAGAGGTTG 1680
561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
1681 ATTGCTTCTGCTGATAGTTGTCATAGCGGGTTGTGGACAGAACTCAAGCCGTAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
1741 GTTCTGAGGCAATTTAAGTCAAGGGCAGGACTTGGAAAGGCCAACACAGCAACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATCTTATCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCCAAAGGCTCTTGTCT 1860
621 GluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGlu 640
1861 GAGAGCTGTTTGTATGATGAAACCTCTTAGTTCGGATTTGATGTCGGAATATATGGA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 660
1921 CAACACTCTGCTCTCGCTCATTTGGGGCACCACAGGCTATGTTGGTTCAGAGAGAGGT 1980

QY 661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680
DB 1981 GGACAACTAAGTCTGAGGCTGTGAGGAGCGGACCTTATGCTCATACTCTTTGATGAAGTG 2040
QY 681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
DB 2041 GAGAAGGCTCATGTTGCTGCTTCAACACCTCTGCTCCCAAGTTTGGATGATGCTGCGATTG 2100
QY 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
DB 2101 ACAGCGGGCAGCAGCAGCAGCTGATTTCCAGAACTCGGTGATAATCATGACATCAAC 2160
QY 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
DB 2161 CTTGCTGCTCAACACACCTCTTTCAGGGCTAACTGGGAAAGTAACTGGAAGTGGCCGG 2220
QY 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
DB 2221 GACTGTGTGATGCGGGAGGTGAGGAAACACTTTCAGACCAGAGCTCTTGAACAGGCTTGAC 2280
QY 761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
DB 2281 GAGATTGTGTGTTGACCCCTTTTCACTGACCACTTGGAGAAAGTAGCTCGGCTTCAA 2340
QY 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
DB 2341 ATGAAAGACGTTGCTGCTCGGCTTGTGTAAGAGAGGAGTTGCTTTGGCAGTCACTGATGCT 2400
QY 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
DB 2401 GCTTTGACTATATCTTTGGCAGAGAGTTATGACCCGCTGTATGCTTAGGCTATAGA 2460
QY 821 ArgTyrMetGluLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
DB 2461 AGATGATGAGAGAGAGAGGTGTGACAGAACTGTCAAAGATGCTTGGCTGAGGAAATC 2520
QY 841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
DB 2521 GATGAAACCTCCACTGTTTACATAGATGACGAGGCTGGTGATCTTGTGTACCGGCTAGAA 2580
QY 861 SerGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
DB 2581 AGTGGAGGCTTAGTGGACGCTTCAACAGGCAAGAGTCAGATGTGCTGATTATATTGCT 2640
QY 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
DB 2641 AACGGGCCAAAGAGAGATGATGACGCTCAGGCGGTGAAGAGATGAGGATCGAGGAAATA 2700
QY 901 GluAspAspAspAsnGluMetIleGluAsp 911
DB 2701 GAAGATGACGATAATGAGGAAATGATCGAGGAT 2733

RESULT 3

US-10-310-154-113

; Sequence 113, Application US/10310154

; Publication No. US20030233670A1

; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; APPLICANT: Chometon, Paul S.

; APPLICANT: Adams, Thomas H

; APPLICANT: Ruff, Thomas G.

; APPLICANT: Agarwal, Ameeta K.

; APPLICANT: Ahrens, Jeffrey E.

; APPLICANT: Ball, James A.

; APPLICANT: Banu, G.

; APPLICANT: Bell, Erin

; APPLICANT: Boddupalli, Raghava

; APPLICANT: Deikman, Jill

; APPLICANT: Deng, Molian

; APPLICANT: Dong, Jinzhao

; APPLICANT: Duff, Stephen M.

; APPLICANT: Galligan, Meghan M.

APPLICANT: Hinchey, Brenda S.
 APPLICANT: Huang, Shihshieh
 APPLICANT: Johnson, G. Richard
 APPLICANT: Jung, Vincent
 APPLICANT: Kretzner, Keith A.
 APPLICANT: Laccetti, Lucille B.
 APPLICANT: Lai, Chao-Qiang
 APPLICANT: Lee, Gary
 APPLICANT: Lin, Jie-Yi
 APPLICANT: Liu, Jingdong
 APPLICANT: Lu, Bin
 APPLICANT: Luethy, Michael M.
 APPLICANT: Lund, Adrian
 APPLICANT: Madson, Linda L.
 APPLICANT: Malloy, Kathleen A.
 APPLICANT: McKiel, Christine L.
 APPLICANT: Miller, Philip W.
 APPLICANT: padnavathi, Manchikanti
 APPLICANT: Parnell, Laurence D.
 APPLICANT: Start, William G.
 APPLICANT: Tennessee, Dan
 APPLICANT: Vidya, K.R.
 APPLICANT: Wang, Haiyun
 APPLICANT: Xin, Zhanguo
 APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping
 APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li

TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants

FILE REFERENCE: 38-15(52796)B
 CURRENT APPLICATION NUMBER: US/10/310,154
 CURRENT FILING DATE: 2002-12-04
 PRIOR APPLICATION NUMBER: 60/337,358
 PRIOR FILING DATE: 2001-12-04
 NUMBER OF SEQ ID NOS: 736
 SEQ ID NO 113
 LENGTH: 3942
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (436)..(1773), (1878)..(2159), (2281)..(2621), (2711)..(2990), (3079)..(3276),
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(3942)
 OTHER INFORMATION: unsure at all n locations
 10-0-310-154-113

Alignment Scores:
 2.35e-291 Length: 3942
 3657.00 Matches:
 78.26% Conservative: 76
 71.23% Mismatches: 60
 80.37% Indels: 177
 12 Gaps: 11

09-812-350-17 (1-911) x US-10-310-154-113 (1-3942)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 436 ATGAATCCGACAACTTCAACCAAGACGAAACGAGCGGATCGTGGGGGGCGCAGAGATT 495
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 496 GCGGTGGAGGCGCCACGCGAGCTCACGCGCGTGCACCTGGCCCGCGAGTGGCTGGC 555
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
 556 GACNAGGGCGGCATCTTGGCGGAGCCCATCACGGGGCGCTGCGGGGGCGGACGAGCGGCC 615

60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 616 GGGGACTCTGTCGAGCGCTGTGAACAACTCGCTCAAGAACTGCGCTCGAGTCCCGC 675
 79 ProProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 676 CCGCCGCGACTCGCTTCGCGCGTCTACGCGCTGATCAAGTTCATCGCGCGCGAGTCC 735
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 736 GCGCAGAAAGACGCGGAGCTCGCACTTCGCGCGTCCAGCACTGCTCTCGCTGCTTC 795
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 796 GAGGACTCGCAGATCTCCGACTCGCTCAAGAGGCGCGCTGTCGCGCGCGGTGGC 855
 139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 856 GCGGAGCTTGAAGACTCGCGCGCGGAGGCGCGCGTGGAGTCCGCGTCCGCGGGAT 915
 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
 916 ACCAACTTCAGCGCTCAAGACATACGCGCGGACCTCGTCAGACAGCGCGGAGTT 975
 179 AspProValIleGlyArgAspGluLeuIleArgArgValValArgIleLeuSerArgArg 198
 976 GACCCCGTTCATCGCGCGGAGGAGATCCGCGCGTCTGCGCATTTCTCTCGCGCGC 1035
 199 ThrLysAsnAsnProValLeuIleGlyLysProGlyValGlyLysThrAlaValGlu 218
 1036 ACCAAGAAATACCCCGTCTCATCGCGGAGCGCGCGTGGCAAGACGCGCGTCTGAG 1095
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 1096 GGCCTCGCGCAGCGATGTTTCGCGCGAGTGCCTCCAGTAACTCTCTCGAGTCCGCTC 1155
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 258
 1156 ATCGCGCTCGACATCGCGCGCTCTCGCGCGCGCAAGTACCGCGCGAGTTCGAGGAG 1215
 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleIeuPheIle 278
 1216 CCGCTCAAGCGCGTCTCAAGAGAGTGAAGAGCGCGGAGGAGGAGTCAATCTCTTCATC 1275
 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
 1276 GACGAGATACCTCTCTCGCGCGCGGAGAGCGAGGAGTTCATGAGCGCGCGCCAC 1335
 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 1336 CTGTTCAAGCAATCTCGCGGAGGAGCAGCTCAGGTGATCGCGCGCCACCACTCGAG 1395
 319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyr 338
 1396 GAGTACCGCAAGTACGTGGAGAGAGCAGCGTTCGAGCGCGGTTCACAGCGGTTC 1455
 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
 1456 GTCGCGGAGCGCGTCCCGCACACCGTCAAGCATTTCTGAGGGGCGCTCAGGAGAGTAC 1515
 359 GluGlyHisLysGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 1516 GAGGGGCACTTGGCGTGGAGTCCAGGACCGCGCTCTGTTGGTGGCGGACAGCTCTCC 1575
 379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
 1576 GCGAGGTACATCATGGTTCGCGCATCTGCTCACAAGACATAGACCTGGTGGAGCGGCC 1635
 399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
 1636 TCGGCATGTGAGGGTGCAGCTCGACAGCCGCGGAGAGATTGATTACCTGGAGAGG 1695
 419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438

1696 AAGAGAAATCCAGCTTGGAGTTGAGCTCCAGCGCTCGAGAGGAGGACAGGCGCAGC 1755
439 LysAlaArgLeuIleGluVal----- 445
1756 AAGCGCGCTGATTGAGGT-CCGTGCTCGAGCACTGAATTTCTCAAAACAAGTCCTC 1814
445 ----- 445
1815 TTGTCTGATGTTCTGTTTGAACCTTTGATTAACCTGCTTAACCTGCTGCGCAATGTT 1874
446 -----ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyr 463
1875 CAGGTCAAGGAAGATTGACGATCTGAGGACAAGCTGACGCCCTGACCATGAAGTAC 1934
464 ArgLysGluLeuGluArgLeuIleArgArgLeuLysGlnLysArgGluGluLeu 483
1935 AGGAGGAGAGAGAGAGAAATGATGAGATCAGGAAGCTGAAGCGCGGAGGAGCTC 1994
484 MetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArg 503
1995 CAGTTCAACCTTCAGAGAGCGCGCGCGGATGACCTGGCCCTGTGGCCGACCTCAAG 2054
504 TyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGlu 523
2055 TACGGCGCCCTCAGGAATCAGCGCTGCTATCTCAAGCTGGAG---AGCGAAACAGGG 2111
524 GluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu----- 539
2112 GAGAACCTGATGCTCACCAGAAACCGTGGCCCTGAACAAATTCAGAGGTATGTTATTAT 2171
539 ----- 539
2172 TCTTGTTCACCGTCACAAAAATTTTCAGAGCAAGTGCGRATTTGCGCATGCTCTCCTA 2231
540 -----ValValSerAr 543
2232 GTAGAGTAGTCTGCAGCGTGTGCTGAAATGCTGTTCTGCTATGGCGAGGTGGTGAGCG 2291
543 GTPThrGlyIleProValThrArgLeuGlyGlnAsnGlyLysGluArgLeuIleGlyLe 563
2292 TTGACGGGTATTCAGTGACCCGCTGGCCAGAACGACAGGAGAGAGCTGTGTGGCT 2351
563 uAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerG1 583
2352 GGCTGACAGGCTTACCAGAGGTGTGTGGCCAGACAGAGGCTGTGTGGCGCTGCAGA 2411
583 uAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLe 603
2412 GCGGTGCTGAGGTGAGGCGCGCTTGGCAGGCCACAAACAGCCCTGCTGCTTCT 2471
603 uPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLe 623
2472 CTTCTCTGGTCCGACTGGCGTGGGAAACTGAGTGGCCAGGCCCTAGCCGACAGCT 2531
623 uPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGluGlnHisSe 643
2532 GTTCACACGACGAACTTCTTGTTCGATCGACATGTGGAGTACATGTGAGCAGCACTC 2591
643 rValSerArgLeuIleGlyAlaProProGly----- 653
2592 GGTGTCGCCCTCATCGGAGCACCACTGG- GTAAGTAGCAGAAAAATGCATGATCTTCG 2650
653 ----- 653
2651 TATTTTAACTCGAATTCGATGAACCTTGTTCGACGTGGCGGACCTGTGACAG 2710
654 -TyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValaArgArgProTyrCy 673
2711 CTACGTGGCCATGAAGAGGTGGGAGCTGACTGAACAAAGTGAGGAGGAGCGCTACAG 2770
673 sValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuGlu 693

2771 CGTGATCTCTTCGAGGAGCTCGAAGGCCCTATGTCGCGTGTCAACACCCCTGCTCCA 2830
693 nValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgense 713
2831 GGTCTCTCGAGCGAGGCTGAGCGATGGCGAAGCGAGGAGCTGAGCTTCAGGAACAC 2890
713 rValIleLeuMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLy 733
2891 CGTGATCATCATGACATCGAACCTCGCGCCGAGAGCACTCTCTGCTGGGATGCTGGGCAA 2950
733 sValThrMetGluValAlaArgAspCysValMet-Arg----- 745
2951 GAATCCATGAAGTCTGCTCGCATCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 3010
745 ----- 745
3011 CTGAATCGTCTGATGACATTTGCTCTGCGAGCGTGAATGCTCAATCGGATCGTTATTGCT 3070
746 -----GluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValV 764
3071 GTGTGACAGGTGAGGAGCACTTCGCGCCCTGAGCTGTGAACCGCTCTCGACGATCGTGA 3130
764 alPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspV 784
3131 TCTTCGATCTCTGTCCACGAGCAGCTGAGGAGGTGCTGCTGCTTCAGATGAGGATG 3190
784 alAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspT 804
3191 TGGCGTCTCTCTTTCGGAAGGGGATGCTCTGCTGTGACCGAGCGCGCATGTCGACA 3250
804 YrIleLeuAlaGluSerTyrAspPro----- 812
3251 TCATCTTGTCTCTCTTACGATCGCTGATGTGACCATCCATGATTTGATTCATCTGAAT 3310
812 ----- 812
3311 TCGTGGTGACACTGATGCTGCTGCTCTTATCTTCTTGTGTGGCTTCAACAACAG 3370
813 ValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSer 832
3371 GTGTATGGCGCGCGCCATCAGAGGTGATCAGAGAGGTGTGAGCAGCTGTG 3430
833 LysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAla 852
3431 AAGATGTGATCCAGGAGGAGATCGACGAGACTGACCGCTCTACATCGAGCGCGCGC 3490
853 Gly-----AspLeuValTyrArgValGlu-----SerGlyGlyLeuValAspAlaSerThr 869
3491 GCGAAGGACAGCTGTGCTACAGGGTGGACCGGCGGCGGTCTGCTGACCGCTGAGAGC 3550
870 GlyLysLysSerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAla 889
3551 GGGATGAAGTCCGACATCTGATCCAGGTCCCCAACAGCTCCACAGGAGCGCTGCG 3610
890 GlnAlaValLysLysMetArgIle-----GluGluIleGluAspAspAsnGluGlu 907
3611 CAGCGCTCAAGAGATGAGATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3667

RESULT 4

US-10-369-493-43734
Sequence 43734, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 43734

LENGTH: 2643

TYPE: DNA

ORGANISM: No. US20030233675Altoc punctiforme
110-369-493-43734

Alignment Scores:

i. No.: 4,68e-181 Length: 2643
2315.00 Matches: 467
Conservative: 172
Mismatch: 53.93%
Indels: 28
Gaps: 11

us-09-812-350-17 (1-911) x us-10-369-493-43734 (1-2643)

2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
34 AATCCGAGCAATTTACCGAAAGACCTGGAGCCCTTGTCTGCTACTCCGAAATGGCC 93
22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAsp 41
94 AAACAGTTTCAGCATCAGCAGATTGAGAGCGCAACATTTGATGCTAGCGCTACTG--GAA 150
42 ProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGlnSer 61
151 CAGGAGGACTC---GCCAGTCTTATTTTCAACAAGCCGGGTAAATGTTCAAAAACATA 207
62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
208 CACGAGCGCACTATCGACTTCATCAACCGTCAGCCCAAGTATCGGGGGTACAGTGGT 267
82 AspIleProAlaSerSerLeuIleLysValIleArgAlaGlnAlaAlaGlnLys 101
268 TCAGTATACATCGACACATTTAGAGAGGCTGCTCGATCGCGCGCAACAAATACCGCAA 327
102 SerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSer 121
328 GAGTTGGTGATGAATATATTTCTATTGAACATTTAATATCTTCCTTTGCCAAAGACGAT 387
122 GlnIle---ArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
388 CGTTTGGTAAAGGGTTGTTCAAGATTGGATGGATGGAATAAACTCCGCAACATC 447
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
448 ATTCAACAGATTCCGAGGAGTCAA-----AAAGTGACAGATCAAAACCGGAAGTTAAA 501
161 PheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAla-----GlyLysLeu 178
502 TATGAGCGGTAGAAATACGGCGCGGATTTAACCAATGGCCATGAGGGCATACTC 561
179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
562 GACCCAGTGTGGCAGATGAAGAAATTCGCGCAGCAGATTCAATCTTCTCTCGCG 621
199 ThrIleAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGlu 218
622 ACTAAATAATACCCCGTGTGATTTGGTGAACCCGGTGTGCGTAAACCGCAATTTAGAA 681
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeu----- 233
682 GGATTAGCGGAGGCAATTTGTCAGTGGTATGTCGCCGAATCATTTGGCGGTGGCATAGCC 741
234 -----ThrAspValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLys 251
742 GGCCGTAGGCAATCGCAACTAATAGCTTTAGACATGGGTGGTAAATTTGCCGGAGCCAA 801
252 TyrglyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGlu 271

802 TACCGGGGAGAAATTTGAAGAACGCTGAAAGCTGTCTTAAAGAAATCCAAAGACACAG 861
272 GlyLysValIleLeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGlu 291
862 GGACAAATCGTCTTGTTCATTGACGAAATTCACACTAGTTGGTGGGGTGCACACGCAA 921
292 GlySerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCys 311
922 GGATCGATGATGCTAGCAACTTGTCTAAGCGATGCTGCTCGCGGGGAACTCGCGCTGT 981
312 IleGlyAlaThrThrLeuGluGlyTyrglyTyrglyValGluLysAspAlaAlaPheGlu 331
982 ATTGTGCGCACCACTAGATCAATACCGCAAGTACATTGAAAAGATGCGGCTTTGAA 1041
332 ArgArgPheGlnGlnValTyrglyValAlaGluProSerValProAspThrIleSerIleLeu 351
1042 CGTCGTTTTACAGAGGTATATGTCGATCAGCCAGTGTGGAAGATACCATCTCTCAATTCTG 1101
352 ArgGlyLeuLysGluLysTyrglyHisHisGlyValArgIleGlnAspArgAlaLeu 371
1102 CCGCGTTTGAAGAGCGCTACGATGTTACACACGCGGTGAAGATTTCTGTAGTCGCTTA 1161
372 IleAsnAlaAlaGlnLeuSerAlaArgTyrglyThrGlyArgHisLeuProAspLysAla 391
1162 GTTGGCGGAGCTACTCTGTCTGCGAGATATATTAGCGATCGTCTTACCCGACAAAGCT 1221
392 IleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGlu 411
1222 ATTGATTTGGTGGATGAAGCTCGCGCTAAACATAAATGGAATTTACTTCCAAACCGAA 1281
412 GluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGlu 431
1282 GAATTAGAGCAATCGATCGCAAAATTTCTGCACTGGAATGGAGCGGCTGTCACTGCA 1341
432 ArgGluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAsp 451
1342 AAAGAAACAGACAGTGTCTCCAGAGAGCGTTTAGAACGGCTGGAGAGAGAACTTTGCAGAG 1401
452 LeuArgAspLysLeuGlnProIleuThrMetLysTyrglyLysGluLysGluArgIleAsp 471
1402 TTAAGAAACCGCAAGATGCGCTCAATGCTCAATGGCAAGCAGAGCAAAATCATCGAT 1461
472 GluIleArgArgLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGlu 491
1462 CGCATTCGCCAAATTAGACAGAGATTGAGCGTGTCAATGTAGAAATTCAGCAAGCCGAA 1521
492 ArgArgTyrglyAspLeuAlaArgAlaAlaAspLeuArgTyrgly----- 505
1522 CGCGATTACGACCTCAACCGCGCAGCAGAAATTTGAATACAGCAAACTCCTCAGCTGCAA 1581
506 ---AlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGlu 524
1582 CGAACACTAAAGAAAGCGCAAGCGCGACTAGCCCAATTCAG---ACTAGCGGTAAATCT 1638
525 AsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTrp 544
1639 -----CTGTTCGCGAAGAGTCAACCGAGCTGACATTCGCGAAATTTCTCCAGTGG 1692
545 ThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAla 564
1693 ACGGTATTCGGTGAGCAGACTGTTGAAATCGGAAATGCAAAACTCTTGCATCTCGAA 1752
565 AspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAla 584
1753 GAAGAGCTACATAAACCGTGTGATTGGTCAGACGCAAGCTGTGAAGAGCTGTTCGGGATGCA 1812
585 IleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPhe 604
1813 ATTCACGTTCCCGTGTGGTGGCAGACCCGCAATCGTCCCATTTGCTAGCTTTATTTTC 1872
605 LeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPhe 624
1873 TTAGCCCGCAGCCGGGGTGGGAAACAGAACTGGCAAAAGCTTTAGCGGAATATCTGTTT 1932

948

316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
 949 ACTTTAGATGAATATCCAAATATATCGAAAGAGATCGGCTTTGGACACGCTTTCCAG 1008
 336 GlnValTyrValAlaGluProSerValProAspThrLysSerLysLeuArgGlyLeuLys 355
 1009 GAAGTTTTAGTGGATGAACCAATGTTATAGATACCAATTTCCATTTCTCCGGGGATTA 1068
 356 GluLysTyrGluGlyHisHisGlyValArgLysGlnAspArgAlaLeuLysAsnAlaAla 375
 1069 GAAGCTATGAAGTACACCGCGGTAATAATTCGCCATAGTCCCTGTGGCGGCGCC 1128
 376 GlnLeuSerAlaArgTyrLysLeuGluGlyHisLeuProAspLysAlaLysLeuVal 395
 1129 ATGTTGTCCCAATCGTACATCAGTATCGTTTCTGCGCGATTAAGCTATTGATTTAGTA 1188
 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluLysLeuVal 415
 1189 GACGAAGCGCGCAAAATTAATAATGGAATCACTCCAAACGAGGGAATTAGATGAA 1248
 416 LeuGluArgLysArgMetGlnLeuGluLysLeuGluHisAlaLeuGluArgGlyAsp 435
 1249 GTTGACCGGAATTTCTCCAACTAGAAATCGAGGTTTATCTTTACACCGGGAATGAT 1308
 436 LysAlaSerLysAlaArgLeuLysGluValArgLysGluLeuAspAspLeuArgAspLys 455
 1309 TCTGCTTCCAGGAGCGCTAGAAAATTCGAGAAAGATGCTGATTTTAAAGAAAGAA 1368
 456 LeuGluProLeuThrMetLysTyrArgLysGluLysGluArgLysLeuValArgArg 475
 1369 CAGTCTAACTTAATGCTCAATGGCAGTGGCAAAACCGGTTATTCATCAAAATTCGTACT 1428
 476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAsp 495
 1429 GTTAAGGAACCATCGACCGAGTGAACCTAGAAATTCACAGCGCCCAACCGGATTCAGAC 1488
 496 LeuAlaArgAlaAlaAspLeuArgTyrGlyValAlaLeuGluValGluSerAlaLeuAla 515
 1489 TACAATAAAGCAGCGAGTTACAGTATGCAATTAATTAATTAATTAATTAATTAATTAAT 1548
 516 GlnLeuGluGlyThrSerSerGluGluAsnVal-----MetLeuThrGluAsn 531
 1549 GCTTTGGAACCAATTCGGGAGCAGCAAACTCTGGCAAAATTCCTCTTACGGGAAGAA 1608
 532 ValGlyProGluHisLeuAlaGluValValSerArgTyrThrGlyLeuProValThrArg 551
 1609 GTTTAGAGTCTGACATGCTGAATTAATCTGAAATGACCGGCAATTCCTCATCAGTAA 1668
 552 LeuGlyGlnAsnGluLysGluArgLeuLysLeuAlaAspArgLeuHisLysArgVal 571
 1669 TTGGTGAATCGGAAAGAAAACCTGCTCCACTTGAAGATGAACACTACACAGCGAGTG 1728
 572 ValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaLeuArgSerArgAlaGly 591
 1729 ATTGGTCAGATGACCGGTAAACCGCGTAGCGAAGCAATCAACGCTCCCGAGCTGT 1788
 592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611
 1789 CTTTCGATCTTAATCGTCCACCGCTAGCTTTATTTCTGGCCCCACACGGGTCCGG 1848
 612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuVal 631
 1849 AAAAATGAGTTAGCGAGGCTTTGGCGAAAAATTTATTCACACGGAAGAGCCCTGGTG 1908
 632 ArgLeuAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuLysAlaPro 651
 1909 CGAATGATATGCTGAATATAGGAAAAACAGCTGTTTCCCGTTTATTCGGGGCCCT 1968
 652 ProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgPro 671
 1969 CCGGGCTATGCGGCTATGAGAGAGGGGACAAATTCACGGAAGCAATTCGCGCGCGGCC 2028

672 TyrCysValLeuLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
 2029 TATTCGGTCATTTCTTTTGACGAGATTGAAGAAGCCCATGGGATGTGTTAACGTCATG 2088
 692 LeuGlnValLeuAspAspGlyArgLeuThrArgGlyGlnGlyArgThrValAspPheArg 711
 2089 CTCCAATCTCGATGATGCGCGTTTAAACGATGCCAAGGCCATGTGTGAGCTTCAAA 2148
 712 AsnSerValLeuLeuMetThrSerLeuLeuGlyAlaGluHisLeuLeuAlaGlyLeuThr 731
 2149 AATACGATTTATCATTTATGACCAAGTAACTGGGCTCCCAATACATTTTGGATGTGGCGGG 2208
 732 GlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPhe 751
 2209 GATGATAGTCGTTATGAAGAAATCGGAGCCGAGTTATGGATGTAAATCGGGAAACCTTC 2268
 752 ArgProGluLeuLeuAsnArgLeuAspGluLeuValPheAspProLeuSerHisAsp 771
 2269 CGCCCAAGATTTCTCAATCGGCTGGATGAACAGATTTATTTCCATGGCTTACAAAAATCC 2328
 772 GlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArg 791
 2329 GAGTTACGATCCATTTGTCCAATTCAAATTCAGTTTGGCTTACCCGTTTGGAGGAACAA 2388
 792 GlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrLysLeuAlaGluSerTyrAsp 811
 2389 AAATTAACCTTTGAAGTTAAACGATGAAGCCCTAGATTTCTGGCTGCGCTGATGAC 2448
 812 ProValTyrGlyAlaArgProLeuArgArgTrpMetGluLysLysValValThrGluLeu 831
 2449 CCGCTTTATGGGCGCGACCTTTAAACAGAGCGCTCCAAAAATACCTAGAAACCGCGATC 2508
 832 SerLysMetValValArgGluGluLeuAspGluAsnSerThrValTyrLysAlaGly 851
 2509 GCCAAGGAATTTTACGGGGGATTAACAACTGGTGAACCATTTGGTGGATGAAC 2568
 852 AlaGlyAspLeuValTyrArgValGluSerGlyGlyLeuVal 865
 2569 GACGAACGCTTCAGTTTACCAGTTTAAAGGGGGATTTAGTC 2610

RESULT 6

US-10-047-260-37
 ; Sequence 37, Application US/10047260
 ; Publication No. US20020164706A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Huang, Lisa
 ; APPLICANT: McClusky, Michael
 ; APPLICANT: Lakoska, Robert
 ; TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA
 ; FILE REFERENCE: CL1715 US NA
 ; CURRENT APPLICATION NUMBER: US/10/047,260
 ; CURRENT FILING DATE: 2002-01-15
 ; PRIOR APPLICATION NUMBER: 60/264,925
 ; PRIOR FILING DATE: 2001-01-30
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 37
 ; LENGTH: 2619
 ; TYPE: DNA
 ; ORGANISM: *Synechocystis* sp. strain PCC6803
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2619)
 US-10-047-260-37

Alignment Scores:
 Pred. No.: 7,74e-179 Length: 2619
 Score: 2288.00 Matches: 460
 Percent Similarity: 71.74% Conservative: 167
 Best Local Similarity: 52.63% Mismatches: 229
 Query Match: 50.29% Indels: 18
 DB: 14 Gaps: 7

09-812-350-17 (1-911) x US-10-047-260-37 (1-2619)

2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
 13 GATCCTTAATAAATTTACGGAGAAAGCTTGGAGGCGATCGCAAAACACCGAGATGCT 72
 22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSerAsp 41
 73 AAACAGCATCGCAACAGCAAAATTTAGACGGAACACCTACTCAGTGCCTACTA--GAA 129
 42 ProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGlnSer 61
 130 CAANAATGGTTCGCCACCGACATCTTTAATAATAGCTGGGGCG-----AGC 174
 62 AlaGluArgVal-----IleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 175 ATTCGCCGAGTTAACGATCAAGTTAATAGCTTTATGCCCCAACGCCAAATAAGTAAT 234
 79 ProProAspIleProAlaSerSerSerLeuIleLysValIleAArgAlaGlnAla 98
 235 CCGAGTGAATCGATTTATTAGCCCGCAGTCTCGATAAATTTGTGGACAATCGGAAATA 294
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 295 GCCAAGTCTAATATGAGACGACGACTATATTCATCGACACTTGATGGCGCTTACGGC 354
 119 GluAspSerGlnIle---ArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
 355 CAAGATGACCGCTGGGCAAAATTTATATCGAGAAATTTGGCTTAACAGAAAATAAGTTG 414
 138 LysSerGluValGluLysLeuArgGlyLysGluGlyLysValGlnSerAlaSerGly 157
 415 GCAGAAATTTATCAAGCAAAATTAGAGGAACCCAA-----AAAGTGACCGATCAAAATCCA 468
 158 AspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAla----- 175
 469 GAGGGCAATACGAATCCCTGAAAATATGGGAGATTTAACGGAATTAGCCCGGAA 528
 176 GlyLysLeuAspProValIleGlyArgAspGluIleArgValValAArgIleLeu 195
 529 GGTAAACTAGATCTGTCTATGCGCGGATGAAGAAGTGGCGCGCACCATTCAGATCCTT 588
 196 SerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
 589 TCCCGCGGCACAAAATAAACCTGTGTTAATTTGGGGAACCGGGGGTTGGTAAACCGCG 648
 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
 649 ATCCGCGAAGTTTAGCCCAAGAATTTAATACCATGACGTACCGGAATCATTCGGGAT 708
 236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGlu 255
 709 CCAAACTAATTTCCCTCGATATGGGGCGGTTAATTTGGCGGGCAAAATATACCGGGGGA 768
 256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIle 275
 769 TTTGAGAAAGACTTAAGCGGTACTTAAAGAAAGTTTACCGAGCCGAGGGGCAATATT 828
 276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
 829 CTCTTTATTGACGAATTCATACCGTTGTGCGCGCTGGGGCCACCCAGAGAGCATGGAT 888
 296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
 889 CGGGGCAACTTTATGAACCCCATGTTAGCCCGGGGTGCTTTGCGTGTATCGGGGCCACC 948
 316 ThrLeuGluGluTyrglySerValGluLysAspAlaAlaPheGluArgArgPheGln 335
 949 ACTTTAGTGAATATCGCAAAATATATGAAAAGATCGGCTTTTGGAAACGACGTTTCAG 1008
 336 GlnValTyrglyAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
 1009 GAAGTTTATGATGAACCAATATGATTAGATACCAATTTCCATTTCCCGGGGATTAATA 1068

356 GluLysTyrglyGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
 1069 GAACCTATGAAGTACACACCGCGTAAATAATGCCGATAGTGCCTCGGTGGTGGCGGCC 1128
 376 GlnLeuSerAlaArgTyrglyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
 1129 ATGTTGTCCAATCGTACATCAGTGATCGTTTCTGCGGATAAAGCTATTGATTAGTA 1188
 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
 1189 GACGAGCAGCGGCCAAATTAATAAATGAAATCACCTCCAAACCCAGAGGAATTAGATGAA 1248
 416 LeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAsp 435
 1249 GTTGACCGGAAATTCCTCAACTAGAAATGAGAGTTGCTGATTTTAAAGAAAGAA 1308
 436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLys 455
 1309 TCTGCTTCCAAGGAGCGCTAGAAAATTTGGAGAAAGTTGCTGATTTTAAAGAAAGAA 1368
 456 LeuGlnProLeuThrMetLysTyrglyArgLysGluLysGluArgIleAspGluIleArgArg 475
 1369 CAGTCTAAACTTAATGCGCAATGCGCAGTCGGAATAACGGTTATTGATCAAAATCGTACT 1428
 476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrglyAsp 495
 1429 GTTAAAGGAACCATCGACAGCTAGAACCTAGAAATTTCAACAGGCCCAACCGGATTACGAC 1488
 496 LeuAlaArgAlaAlaAspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAla 515
 1489 TACAATAAGCAGCGGAGTTACAGTATGCAATTAACCTGATTTACAGCGGCAAGTGGAA 1548
 516 GlnLeuGluGlyThrSerSerGluGluAsnVal-----MetLeuThrGluAsn 531
 1549 GCTTTGGAACCCCAATTTGGCGGAGCAACAACTCTGGCAAAATCCCTCTTACGGGAAGAA 1608
 532 ValGlyProGluHisIleAlaGluValValSerArgTyrglyIleProValThrArg 551
 1609 GTTTTAGAGTCTGACATGCTGAAATTTCTCGAAATGACCGGCATTCCTCATCAGTAA 1668
 552 LeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgVal 571
 1669 TTGTTGGAATTCGGAAGAAAACCTGCTCCATTTTGGAAAGATGAATCACACGCGAGTG 1728
 572 ValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGly 591
 1729 ATTGTCAGATGACGCGTTAACCGCGGTAGCGAAGCCATTCACGCTCCCGAGCTGT 1788
 592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611
 1789 CTTTCCGATCCTAATCGTCCACCGCTAGCTTTTATTTTCTGGGCCCCACAGGGTGGG 1848
 612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuVal 631
 1849 AAACCTGAGTTAGCGAAGGCTTTGGCGAATAATTTATTCGACGGAAGAACCCCTGGTG 1908
 632 ArgIleAspMetSerGluTyrglyMetGluGlnHisSerValSerArgLeuIleGlyAlaPro 651
 1909 CGGATTGATATGCTGAATATATGGAATAACACGCTGTTTCCGTTTAAATGGGGGCCCT 1968
 652 ProGlyTyrglyValGlyHisGluGluGlyGlnLeuThrGluAlaValAArgArgArgPro 671
 1969 CCGGGCTATGTGGCTATGAGAAGGGGACAAATTCACGGAAGCAATTCGCGCGCGGCC 2028
 672 TyrglyValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
 2029 TATTCGCTCATCTTTTACGAGATTTGAAAGGCCCATGGGGATGTGTGTAACTCATG 2088
 692 LeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArg 711
 2089 CTCCAATCTCGATGATGCGCTTTTAAACCGATGCCAAGGCCATGTGTGTGACTTCAA 2148

844 GGTAACTAGATCTCTGCTATTTGGCCGGATGAGAAAGTCGGCGGCACCAATTCAGATCTTT 903
 196 SerArgArgThrLysAsnAsnProValLeuLeuLeuGlyGluProGlyValGlyLysThrAla 215
 904 TCCCGCGGCACAAAATAAACCCTGTGTTAAATGGGGAACCGGGGGTGGTAAACGGCG 963
 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
 964 ATCCGCGAAGGTTTAGCCCAAGAAATATTAACCATGACGTACCGGAATCAATTCGGGGAT 1023
 236 ValArgLeuLeuSerLeuAspMetGlyAlaLeuValAlaGlyValLysThrArgGlyGlu 255
 1024 CGCAAACTAATTTCCCTCATATGGGGGTTAAATTCGGGGGCAAAATACCGGGGGGAA 1083
 256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluLysValIle 275
 1084 TTGGAAGAAAGACTTAAAGCGGTACTTAAAGAGTTTACCGACAGCCAGGGGCAAAATATT 1143
 276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluLysSerMetAsp 295
 1144 CTCCTTATTGAGAAATTCATACCGTTGTCGGCGCTGGGGCCACCAAGAGGCCATGGAT 1203
 296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
 1204 CGGGGCAACTTATTGAACCCCATGTTAGCCCGGGTGTCTTTCGTTGTATCGGGGCCACC 1263
 316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
 1264 ACTTTAGATGAATATCGCAATATATCGAAAGAGTTCGGGCTTTGGACGACGCTTTCCAG 1323
 336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
 1324 GAAGTTTGTAGTGAGTGAACCCCATGTTAGTATACCATTTCTCCGGGGATTAATAA 1383
 356 GluLysTyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
 1384 GAACGCTATGAAGTACACACCGCGCTGTAATAATTCGCGATAGTCCCTGTTAGCGCGGCC 1443
 376 GlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
 1444 ARGTTGTCCAAATCGTTACATCATCATCGTTTCTGCGGATTAAGTATTTAGTATTA 1503
 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAsnAsn 415
 1504 GACGAGCAGCGGCCAAATTAATAATGGAATACCTCCAAACACAGAGAAATTAGTGA 1563
 416 LeuGluArgLysArgMetGlnLeuGluLeuGluLeuHisAlaLeuGluArgGluLysAsp 435
 1564 GTTGACCGGAAATTCCTCAACTAGAAATGGAGCGTTTATCTTTACACCGGGAATGAT 1623
 436 LysAlaSerLysAlaArgLeuLeuGluValArgLysGluLeuAspAspLeuArgAspLys 455
 1624 TCTGTTTCCAGAGCGGCTGAAATAATGGAGAAAGATGGCTGATTTTAAAGAGAA 1683
 456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
 1684 CAGTCTAAACTTAATGGCCATGGCAGTCGGAATAACGGTTATTTGATCAATTCGTACT 1743
 476 LeuLysGlnLysArgGluGluMetPheSerLeuGlnGluAlaGluArgArgTyrAsp 495
 1744 GTTAAGGAACCATCGACAGGTGAACCTAGAAATTCACAGGCCCAACGGGATTCAGC 1803
 496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAla 515
 1804 TACAATAAGCAGCGGAGTTACAGTATGGCAATTAAGTATTTACAGCCGCAAGTGGAA 1863
 516 GlnLeuGluGlyThrSerSerGluGluAsnVal-----MetLeuThrGluAsn 531
 1864 GCTTTGGAAACCAATTTGGCGAGCAACAAACCTCTGCAATCTCTTACGGGAAGAA 1923
 532 ValGlyProGluHisIleAlaGluValValSerArgThrGlyIleProValThrArg 551

1924 GTTTTAGAGTCTGACATTTGCTGAAATTTATCTCGAAATGGACCGCATTCCTCATCAGTAAA 1983
 552 LeuGlyGlnAsnGlnLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysAlaVal 571
 1984 TTGTTGGAATCGGAAGAAAGAACTGCTCCACTTGGAAAGTGAACACTACACAGCCGAGTG 2043
 572 ValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGly 591
 2044 ATTGGTCAGGATGAAGCGGTAAACCGCGTAGCGCAAGCCATTCAAGCTCCCGAGCTGGT 2103
 592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611
 2104 CTTTCCCATCTATCTGCTCCACCGTAGCTTTATTTTCTGGGCCCCACAGGGGTCGGG 2163
 612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuVal 631
 2164 AAAACTGAGTTAGCGAAGCTTTGGCGAAAAATTTATTCGACACGGAAGAGCCCTGGTG 2223
 632 ArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaPro 651
 2224 CGGATTGATATGTCGAATATATGGAATAACACGCTGTTTCCGTTTAAATGGGGGCCCT 2283
 652 ProGlyTyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgPro 671
 2284 CCGGGCTATGTGGCTATGGAAGAGGGGCAATTCAGGAAGCAATTCGCCCGCGGCC 2343
 672 TyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
 2344 TATTCGGTCACTCTTTTACGAGATTAAGAAAGCCCATCGGGATGTGTTTAAACGTCATG 2403
 692 LeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArg 711
 2404 CTCCAATCTCGATGATGCCCTTTTAAACCGATGCCCCAAGGCCCATGTGGTGGACTTCAAA 2463
 712 AsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThr 731
 2464 AATACGATTATCATATGACCATTAACCTGGGCTCCCATACATTTGGATGTGGCGGG 2523
 732 GlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPhe 751
 2524 GATGATAGTCTGTTATGAAGAAATGCGAGCGAGTTATGATGTAATGCGGAAACTTC 2583
 752 ArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAsp 771
 2584 CGCCCAAGATTTCTCAATCGGGGATGAACGATTTTTCATGGCTTACAAAATCC 2643
 772 GlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArg 791
 2644 GAGTTAGATCCATTGTCCAAAATTCAAATTCAGTCTTTGGCTACCGTTTGGAGGAACA 2703
 792 GlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAsp 811
 2704 AAATTAACCTTTGAAGTTAAACGATAAGCCCTAGATTTTCTGGCTCGCGCTGATGAC 2763
 812 ProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeu 831
 2764 CCGTTTATGGGCCCCGACCTTTAAACAGCGCGCTCAAAAATACCTAGAAACGGCGATC 2823
 832 SerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGly 851
 2824 GCCAAGGAATTTTACGGGGGATTAACAACCTGGTGAGACCATTTGTGTGTGATGAACC 2883
 852 AlaGlyAspLeuValTyrArgValGluSerGlyGlyLeuVal 865
 2884 GACGAACCGCTCAGTTTATACCATGTTTAAAGGGGATTTAGTC 2925

RESULT 8

US-10-156-761-7220

; Sequence 7220, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 EQ ID NO 7220
 LENGTH: 2637

SEQUENCE: 2,436-178 Length: 2637
 Matches: 460
 Conservativity: 160
 Mismatches: 231
 Indels: 18
 Gaps: 8

SEQUENCE: 10-156-761-7220 (1-2637)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 1 ATGGACATGAACCGCTCACCAGAGTCCAGGAGCGCTCCAGGAGGCCAGACCGCG 60
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 61 GCCGCGCGCATGGGGGACACAGCGTGCAGCGGGAACACCTGCTGCTCGACTTCTCGAT 120
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 121 CAGGAGGACGCTGTATCCCGGTGTCACAGCGCGGACCGAG-----CCGAG 174
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro--- 79
 175 GAATGCGCGCGCGCTGCGGAGGAACTCTCCACCGCGCGGAGGCGCGCGCGCG 234
 80 -----ProAspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGln 97
 235 GCGCACCCCGCGCGCTTCTGTCACCGAGCGCTCGCGGCTGCTCGACGCGCGCGAG 294
 98 AlaAlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeu 117
 295 CGGAGGAGCCAAACCGCTCAGGACAGTACGTGCTCGTGAGACACTCTCTGCTCGCGCTG 354
 118 LeuGluAspSerGlnIleArgAsp-----LeuLeuAsnGluValGlyValAlaThr 134
 355 GCTGAGGAGAGCTCTTCGACCGCGCGCGGCTACTGCTCAACAGCGCGCGCATCACCGAG 414
 135 AlaArgValLysSerGluValGluLysLeuArgGlyLysGluGlyLysValGluSer 154
 415 GACTGTTCTGAGCGCGCTCACCAGGCTCGCGGCAACACAG-----CGGGTCACCTCC 468
 155 AlaSerGlyAspThrAsnProGlnAlaLeuLysThrTyrglyArgAspLeuValGluGln 174
 469 GCCAACCCGAGTGGCTACGAGGCTCTGAGAGTACCGCGCGGACCTTGTCTCTCGAG 528
 175 Ala-----GlyLysLeuAspProValIleGlyArgAspGluIleArgArgValVal 192
 529 GCTCGGTCCGCGCGCTGGACCGGTCATCGCGCGGACGCGGAGATCCGCGCGCTGACC 588

193 ArgIleLeuSerArgArgThrLysAsnProValLeuIleGlyGluProGlyValGly 212
 589 CAGATCTCTAGCCGCAAGACCAAGAACACCCCTCTCATCGCGGACCCCGCGCTCGGC 648
 213 LysThrAlaValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSer 232
 649 AAGACCGCCATCGTCGAGGCGCTGGCCAGCGCATGTTTCGCGCGGACGTCCTCCGAGGGC 708
 233 LeuThrAspValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyr 252
 709 CTGCGCGCAAGACGCGTGTTCGCTCGACATGCGTCTCTGCTGCGCGCGCGCAAGTAC 768
 253 ArgGlyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGly 272
 769 CGCGGGAGTTCGAGGACCGCTCAGGCGCTGCTCAGCGAGGTCAAGGCCCGCGAGGGG 828
 273 LysValIleLeuPheIleAspGluIleHisLeuValLeuGlyLysThrGluGly 292
 829 CGGATCTCTCTTCTGTCGACGAACTCCACACCGTCTGAGGCGCGCGCGCGGAGGG 888
 293 SerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIle 312
 889 GCCATGACCGCGGCAACATGCTCAAGCGCATGCTCGCGCGCGCGGCAATCCACATGATC 948
 313 GlyAlaThrThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArg 332
 949 GCGCGCACCATCTCGACGAGTACCGACAGCATCGAGAGGACGCGCGCTCGACCGC 1008
 333 ArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArg 352
 1009 CGCTTCAGCAGGTCCTGCTCGACGAGCGGAGCGTGGAGACACCATCTCTCCATCTCTGCGC 1068
 353 GlyLeuLysGluLysTyrGluGlyHisHisGlyValArgIleGluAspArgAlaLeuIle 372
 1069 GGAATCGCGGAGCGCTCGAGGTCTTCCAGCGGTGAGATCCAGGACACCGCGCTGTGTC 1128
 373 AsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIle 392
 1129 TCCGCGCGCACCTCAGCCACCGCTACATCAGCATCGGTCTGCTGCTCGCGGACAGGCCATC 1188
 393 AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu 412
 1189 GACCTCGTCGAGGAGCGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1248
 413 IleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArg 432
 1249 CTCGAGCATCACCAGCGCGCTCACCCTGAGATCGAGAGCGCGCGCTGCTGCTGCTGCTGCTG 1308
 433 GluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeu 452
 1309 GAGAGCGACCCCGCAGCAAGACCGCTGAGGAGTGGCGAGGAGTGGCGAGGAGTGGCGAGCTG 1368
 453 ArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGlu 472
 1369 CCGCGAGGCGCGACCGCAACACCGCTGAGGAGCGCGGCGGAGCGCGCTGCTGCTGCTGCTGCTG 1428
 473 IleArgArgLeuLysGlnLysArgGluLeuMetPheSerLeuGlnGluAlaGluArg 492
 1429 GTGAGGAACTCGCGGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGG 1488
 493 ArgTyrAspLeuAlaArgAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSer 512
 1489 GCCTACGACCTCAACCGCGCGCGGAACTCCGCTACGCGCTGAGCGCGCTGAGGAGTGGAGGAG 1548
 513 AlaIleAlaGlnLeuGlu-----GlyThrSerSerGluGluAsnValMetLeuThr 529
 1549 CGACTCGCGCGGAGGAGGAGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1608
 530 GluAsnValGlyProGluHisIleAlaGlnValValSerArgTyrThrGlyIleProVal 549
 1609 GAGGTCTGTCAGGAGGAGGAGTCCCGGATGCTGCGCGCTGAGCGCGGATCCCGCTGCTGCTG 1668
 550 ThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLys 569

1669 GCCCGCTCCAGGAGGCGAAGCGAGAACTGTCGCCCTCCAGAGAGATCTCCGCGCG 1728
570 ArgValValGlyGlnAenGlnAlaValSerGluAlaLleLeuArgSerArg 589
1729 CGCGTCATCGCCAGGAGGAGCGTCAAACTCGTCCAGGAGCCATCATCCGCGCGC 1788
590 AlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGly 609
1789 TCCGCGATCCGCGACCTCGCGCGCCCATCGGCTGTTCTATCTCTCCGCGCCACCGCGC 1848
610 ValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspGluAsnLeu 629
1849 GTCCGGAAGACCGAGCTGCCAACACCTCGCCCGGACTGTTCGATCCGAGGAGAAC 1908
630 LeuValArgLleAspMetSerGluThrMetGluGlnHisSerValSerArgLeuLleGly 649
1909 ATGTCCTCGCTCGCATGAGCGAATACAGGAGCGGCACACCGTCAAGCGGCTCATGGC 1968
650 AlaProGlyGlyThrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArg 669
1969 GCACCGCGCGATACGTCGCTACGAGGAGGCGCGCAGCTCACCGAGGCGGTACGCGCGC 2028
670 ArgProThrCysValLleLeuPheAspGluValGluLysAlaHisValAlaValPheAsn 689
2029 AAGCGGTACTCGGTGCTGTTTCGAGAGATCGAGAGCGGCACACCGATGTTCTCAAC 2088
690 ThrLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAsp 709
2089 ACCCTGCTCAGATCTTCGACGAGCGCGCATCACCGACCGCCGCGCGCACCGTCGAC 2148
710 PheArgAsnSerValLleLeuMetThrSerAsnLeuGlyAlaGluHisLeuAlaGly 729
2149 TTCCGACACCGGTGATCATGACGCTCAACATCGGCTCCGAGCACCTTCTCGACGCG 2208
730 LeuThr-----GlyLysValThrMetGluValAlaArgAspCysValMetArgGluVal 747
2209 GCCACCGCGGAAGTGATGAGTCAAGCCCGAC---GCCCGCGCCCTGCTGATGCGGAGCTG 2265
748 ArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluLleValValPheAspPro 767
2266 CGCGGCGACTTCGCGCGGAGTTCCTCAACCGCGTCAAGACATCGTGTGTTCAACCG 2325
768 LeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArg 787
2326 CTGGGTGAGCGGAGATCGAGCGATCGTGGAACTGCTGACGAGCTCGCGGAGCGG 2385
788 LeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrlleLeuAla 807
2386 CTCCGCGAAGCGCGCATCACCGTCAACTCACCGACGCGCGCGGCGGAGTGTGCGCCAC 2445
808 GluSerTyrlleAspProValTyrlleAlaArgProIleArgArgTrpMetGluLysVal 827
2446 CAGGCTACGACCGGTGTACCGGCGCGCGCGCTGCGCGCTGCTATCTTCCACGAGGTC 2505
828 ValThrGluLeuSerLysMetValValArgGluGluLleAspGluAsnSerThrValTyrl 847
2506 GAGACACTGTCGAGCGCGCTGCTGCGCGGCGAGCTCCAGGACGCGCGAGCTGTCGCG 2565
848 IleAspAlaGlyAlaGlyAspLeuVal 856
2566 GTCCGCGCGGAGCACGAGAGTGGTG 2592

SULT 9

Sequence 1, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:

Pred. No.:	4 6e-174	Length:	9025608
Score:	2282.00	Matches:	475
Percent Similarity:	68.69%	Conservative:	170
Best Local Similarity:	50.59%	Mismatches:	258
Query Match:	50.15%	Indels:	37
DB:	15	Gaps:	12

US-09-812-350-17 (1-911) x US-10-156-761-1 (1-9025608)

QY 1 MetAenProGluLysPheThrHisLysThrAsnGluThrIleAlaThrHisGluLeu 20
Db 8634077 ATGACATGAACCGTCTCACCAGAGTCCAGGAGGCCCTCCAGGAGGCCACACCGCG 8634136
QY 21 AlaValAenAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 8634137 GCCGCGCGCATGCGGGGCACACCGAGGTGCGAGCGGAAACACCTGTGCTCGCACTTCTCGAT 8634196
QY 41 AspProThrGlyLlePheProGlnAlaIleSerSerAlaGlyGlyGluAenAlaAaGln 60
Db 8634197 CAGGAGGAGGTCTGATCCCGGTTCTGCAACAGCGCCGCGACCGAG-----CGAAG 8634250
QY 61 SerAlaGluArgValIleAenGlnAlaLeuLysLysLeuProSerGlnSerProPro--- 79
Db 8634251 GAACTGCGCGCGCGGTGCGGAGGAACTCTCCACCGCGCGAGCGCGCGCGCGCG 8634310
QY 80 -----ProAspAspIleProAlaSerSerSerLeuLysValIleArgAlaGln 97
Db 8634311 CGCGCACCGCGCGCGGTCTTCGTCACCGCGCGCTCCCGCGGTGCTCGACCGCGCGCG 8634370
QY 98 AlaAlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeu 117
Db 8634371 CGCGAGGCCAAACCGCTCAAGGAGGAGTGTGCTGTCGAGCACCTCTCTGCTCGCGCTG 8634430
QY 118 LeuGluAspSerGlnIleArgAsp-----LeuLeuAenGluValGlyValAlaThr 134
Db 8634431 GCTGAGGAGAGCTCTTCGACCGCGCGCGGTACTGCTCAACAGCGCGCATCACCGAG 8634490
QY 135 AlaArgValLysSerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSer 154
Db 8634491 GACTCGTCTCGAGCGCGCTCACCGCGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 8634544
QY 155 AlaserglyAspThrAenPheGlnAlaLeuLysThrTyrlleArgAspLeuValGluGln 174
Db 8634545 GCCAACCGCGAGTGGCTTACGAGGCTCTGGAGAGTACCGCGCGAGCTGTGCTCTCGAG 8634604
QY 175 Ala-----GlyLysLeuAspProValIleGlyArgAspGluGluArgValVal 192
Db 8634605 GCTCGGTCCGCGCGGTGAGCGCGGTGAGCGCGGTGAGCGCGGTGAGCGCGGTGAGCG 8634664
QY 193 ArgIleLeuSerArgThrLysAenAenProValLeuIleGlyLysProGlyValGly 212

8634665 CAGATCTCAGCGCGAAGACCAACACCCCGTCTCTCATCGCGGACCCCGCGGTCTCGGC 8634724
 213 LysThrAlaValValGluGluGluAlaGlnArgIleValLysGlyAspValProAsnSer 232
 8634725 AAGACCGCATCTGTCAGGCGCTGGCCAGCGCATGTTTCGGCGGACGCTCCCGGAGGC 8634784
 233 LeuThrAspValArgLeuLeuSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyr 252
 8634785 CTGCGCGCAAGAAGCGGTTCGCGCTTCGACATGGGTCTCTCTGTCGCGCGCCAGTAC 8634844
 253 ArgGlyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGly 272
 8634845 CGCGGGAGTTCAGGAACGCTCAAGGCGGTCTCAGCGAGGTCAAGGCCCGCAGGGG 8634904
 273 LysValIleLeuPheLeuAspGluLeuHisLeuValLeuGluAlaGlyLysThrGluGly 292
 8634905 CGGATCTCTCTCTCTGTCAGCAATCTCCACCGTCTGAGGAGCGCGCGCGCAAGG 8634964
 293 SerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIle 312
 8634965 GCATGAGCGCGGCAACATGCTCAAGCGCATGCTCGCGCGCGCGCAACTCCACATGATC 8635024
 313 GlyAlaThrThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArg 332
 8635025 GCGGCCACCCCTCGACGAGTACCGAAGCACATCGAGAGGACCGCGCTCGAACGC 8635084
 333 ArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArg 352
 8635085 CGCTTCCAGCAGTCTCTGTCAGCAGCGCGAGCGTGGAGGACCATCTCCATCTCTGCGC 8635144
 353 GlyLeuLysGluLysTyrGluGlyHisGlyValAlaGlnGlnAspArgAlaLeuIle 372
 8635145 GGACTGCGGCAAGCGCTCGAGTCTTCACGCGGTGAAGATCCAGGACCGCGCGTGTGTC 8635204
 373 AsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIle 392
 8635205 TCCGCGGCACCCCTCAGCACCGCTACATCACCAGTCTGTTCTTCTGCCGACAGGCATC 8635264
 393 AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu 412
 8635265 GACCTCTGTCAGAGCGTCTGCGCGGTGCGTACCGAGATCGATCGATCGCGCGGAA 8635324
 413 IleAspAsnLeuGluArgLysArgMetGlnLeuGluLeuGluHisAlaLeuGluArg 432
 8635325 CTCGACGAGATCACCGCGCGTCAACCGCTGAGATCGAGGAGCGCGCTCTCAAG 8635384
 433 GluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeu 452
 8635385 GAGACGACACCCCGCAGCAAGACCGCTGAGGAGCTGCGCAGGAACTGGCGGACCTG 8635444
 453 ArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGlu 472
 8635445 CGCGGAGCGCGGACGCCAACACACCGCCCGTGGAGGCGCGAACCGGAGGATCCGCGC 8635504
 473 IleArgArgLeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArg 492
 8635505 GTGCAGGAATCTGCCAGGAACTGACAGCTCGCCCGCCAGCGCGGAGGAGCGCAACGC 8635564
 493 ArgTyrAspLeuAlaArgAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSer 512
 8635565 GCCTACGACCTCAACCGCGCGGCACTCGCTACCGCTCCAGGACCTGGAGCGC 8635624
 513 AlaIleAlaGlnLeuGlu-----GlyThrSerSerGluGluAsnValMetLeuThr 529
 8635625 CGACTCGCGCGGAGGAGGAGCAACTCGCGCGCCCAAGCAAGGAGGAGCAACCGCTCTGCGC 8635684
 530 GluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProVal 549
 8635685 GAGTCTGTCACCGAGGAGGAGATCGCCGAGATCGTCCCGCTTGGACCGGCACTCCCGTC 8635744
 550 ThrArgLeuGlyGlnAengluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLys 569
 8635745 GCGCGCTCCAGGAGGCGCAACCGCAAACTGTCGCGCTCTCGACGAGATCTCTCGCGCGAG 8635804

QY 570 ArgValValGlyGlnAengluAlaValAsnAlaValSerGluAlaIleLeuArgSerArg 589
 Db 8635805 CGCTGTCAGCGGAGGAGGAGCGCTCAAACTCGTCACCGAGCGCATCATCGCGCGCGC 8635864
 QY 590 AlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGly 609
 Db 8635865 TCCGCGCATCGCGGACCTCGCGCGCCCATCGGCTCTGTTCTCTCTCTCGCGCGCGC 8635924
 QY 610 ValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeu 629
 Db 8635925 GTCCGGAAGACCGAGCTGGCCAGACCTCGCGCGGACTCTGTTCGACTCCGAGGAGAAC 8635984
 QY 630 LeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuLeuGly 649
 Db 8635985 ATGCTCGCTCTGACATGAGCAATACAGAGCGGCGACCGCTCAGCGCGCTCATGGC 8636044
 QY 650 AlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArg 669
 Db 8636045 GCACCGCGCGATACGTCTGCTACGAGGAGCGCGGCGGCTCACCGAGGCGCTACGCGC 8636104
 QY 670 ArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsn 689
 Db 8636105 AAGCGTACTCGTCTGCTGCTTCGACGAGATCGAGAGCGCGCACCGATGTCTTCAAC 8636164
 QY 690 ThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAsp 709
 Db 8636165 ACCTGCTGAGATCTCTGACGAGCGCGCGCATCACCGAGCGCGCGCGCTCTCGAC 8636224
 QY 710 PheArgAsnSerValIleLeuMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGly 729
 Db 8636225 TTCGCAACACCGCTGATCATGATCGTCCAACTCATCGCTCCGAGCGCGCGCGCTCTCG 8636284
 QY 730 LeuThr-----GlyLysValThrMetGluValAlaArgAspCysValMetArgGluVal 747
 Db 8636285 GCACCGCGCGAGGTGAGATCAAGCCGAC---GCCCGCGCGCTGCTGATGGCGGAGCTG 8636341
 QY 748 ArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspPro 767
 Db 8636342 CGCGGCGACTTCGCGCGGAGTCTTCAACCGGTGCGAGCATGCTGCTGTCTTCAACCG 8636401
 QY 768 LeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArg 787
 Db 8636402 CTGGGTGAGCGGCGAGATCGAGCGGATCGTGGAACTGCACTTCGACGAGCTGGCGGCG 8636461
 QY 788 LeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaLeuAspTyrIleLeuAla 807
 Db 8636462 CTCGCGAAGCGCGCATCACCTCGAACTCAACCGCGCGCGGCGCGGAGTGTATCGCC 8636521
 QY 808 GluSerTyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysVal 827
 Db 8636522 CAGGCTACGACCGCGGTGTACGCGCGCGCGCGCTGCGCGCTGCTATCTCCACGAGTCT 8636581
 QY 828 ValThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyr 847
 Db 8636582 GAGACATCTGTCGAGCGCGCGCTGCTGCGCGCGCGCGCTGCGCGCTGCTATCTCCAC 8636641
 QY 848 IleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSer----- 861
 Db 8636642 GTCGACCGCGGACGAGAGTGTGTGTGT-CACCTTACGACCGCGCGGAGCGTCAAGGG 8636700
 QY 862 -----GlyGlyLeu-----ValAspAlaSerThr-----GlyLys 871
 Db 8636701 AGCGTGGCGGCGATGACGACGATGACGCGGAGCGCGCTGCTGCTGCTGCTGCTGCT 8636760
 QY 872 LysSerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAla 891
 Db 8636761 ACCAACCGGCTGCG 8636820
 QY 892 ValLysLysMetArgIleGluGluIleGluAspAspAsnGluGluMetIleGlu 910
 Db 8636821 CTACCGTGTGATG-----GTGACCGCGGAGCGACGAGCTTTCACCGAGGTCTCGCG 8636871

RESULT 10

Sequence 44175, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 44175

LENGTH: 2571

TYPE: DNA

ORGANISM: Rhodospseudomonas palustris

-10-369-493-44175

Alignment Scores:

Seq. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
1	5.52e-177	2571	458	159	234
2	2264.50	71.58%	53.13%	49.77%	11
3	71.58%	53.13%	49.77%	11	7

-09-812-350-17 (1-911) x US-10-369-493-44175 (1-2571)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGAACGTTGAATAATATACCGAAGTGTGGCGGCTTCATCCAGTCAGCGCAATCGCTG 60
21 AlavalAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
61 GCGATGCGCGAGGGGCCATCAGCAGTCTCCGCGCTGCACATTCGAAAGTTCTGCTCGAC 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAlaAlaGln 60
121 GATTCGGAAGGGCTCGCGGTGTGTCATCGACCGCGCGCGCGCAATTCGCGTGGC--- 177
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSer---ProPro 79
178 ---ATCCTGAAGCGACCGAGAGGCGCTCGGCAGATGCCGAAGTGTCCGCTCGGC 234
80 ProAspPheProAlaSerSerLeuIleLysValIleArgAlaGlnAlaAla 99
235 GCGGCGCAAGCTATCTGGCCCGCGGACCGCGCGCGCTTCGACGCTCGGAGAGCG 294
100 GlnLysSerArgGlyAspThrHisLysAlaValAspGlnLeuIleMetGlyLeuLeu--- 118
295 GCGGAAAGCGGCGGACGCTTCGACCGTCCGCGGCTCTCTCGCGCTGTGCTC 354
119 ---GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
355 GATAAGCAGACAGCGCGCTACGCTGCTCACCAGGCGCGGCTACCCCGCAGAACCTC 414
138 LysSerGluValGlnLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGly 157
415 AATGCGGCCATCAACGCCCTGCGC-----AAGGCGGTACCGCGGATTCGCGAGCGGC 468
158 AspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla----- 175
469 GAGAACGCTATGACCGCGTGAAGAAATACGCCCGGACCTCACCGAGCGCGACCGGAC 528
176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeu 195
529 GGCAAGCTCGACCGCGTGTGCGCGCGAGGAAATTCGCGCGACCAATCCAGGTTCTG 588

QY 196 SerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
Db 589 TCGCGCGCGACCAAGAACACCCGGTGTCTGATCGCGCAACCCGGCGTCTGTAAGACCGCG 648
QY 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
Db 649 ATCGTCGAAGGCTTGGCGCTGCGCATCTCAACGGCGAGCTGCCCGAGAGCCTGAAGGAC 708
QY 236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGlu 255
Db 709 AAGAAGCTGCTGCGCTCGACATGGCGCGCTGATTCGGGTGCGAAGTATCGCGCGAG 768
QY 256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluLysValIle 275
Db 769 TTGCAAGACGCGCTGAAGCGCTGCTCAACGAGGTCACCGCGCGCGCGCGCATATC 828
QY 276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
Db 829 CTGTTTCATCGACGATGCACACCTGTTGCGCGCGCGCAAGCGCGCGCGATGAC 888
QY 296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
Db 889 GCGTCGAATCTCTCAAGCCCGCGCTTGCCTGCGCGAGTGCATGCTGCGCGCGGAC 948
QY 316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
Db 949 AGCTTCGATGATATCGACACCTCGAGAGGACCGCGCTGCGCGCGCGCTTCAG 1008
QY 336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
Db 1009 CCGTGTGTTGCTGCGCGCGCGCTGCGAGCACCGTCTGATCTCTGCGCGCGCTGAAG 1068
QY 356 GlnLysTyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
Db 1069 GACAATACGACGACACCGCGCGCTGCGCATCGCGACTCGCGCGCTGCTGCGCGCGCTC 1128
QY 376 GlnLeuSerAlaAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
Db 1129 AGCTTTTCCACCGCTACATCCGACCGCTCTCTCCGCAAGGCGATCGACTGATG 1188
QY 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
Db 1189 GACGAGCGCGCGCGCGCTGAGATGCGAGTCTGACTCCAAAGCCGGAAGAACTCGACTCG 1248
QY 416 LeuGluArgLysArgMetGlnLeuGluIleGlnLeuHisAlaLeuGluArgGlyLysAsp 435
Db 1249 ATGACCGCGAGATCGTGGCGCTGAAGATCGAGCGAGCGCTTGAAGAGGAACCGCAT 1308
QY 436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLys 455
Db 1309 CCGGCTCGAAGCGCGCTGTTGACGCTGGAGAGGAGCTTCCGATCTCGAGAGAGAG 1368
QY 456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
Db 1369 TCGCGCGGCTTGAAGCGCGCTTGGAGTCCGAGAGAAACAAGCTGTCGATGCCAAG 1428
QY 476 LeuLysGlnLysArgGluLeuMetPheSerLeuGlnAlaGluArgArgTyrAsp 495
Db 1429 CTGAAGAGCGAATCGACGCTCTCGCGATTTGAATCTGCCAACCGCGCGCGCGGTGAG 1488
QY 496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAla 515
Db 1489 TATCAGCGCGCGCGAGCTCGCTATGGCGGATTCGAGGCTTGAAGAGAGATCGCT 1548
QY 516 GlnLeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGlu 535
Db 1549 GAGATCGAA---GCCACGAGAACTCGCGCGCGATGTTGGAGGAGCGCGTCCCGCAAT 1605
QY 536 HistLeuAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsn 555
Db 1606 CATCTCCCGAGGTGTGCGCTGACCGCGCTGCGCGCGTCCGAGATGCTCGAAGGC 1665

556 GluLysGluArgLeuLeuLeuAlaAspArgLeuHisLysArgValValGlyGlnAsn 575
 1666 GAGAGGAAAGCTCGTGGTGAAGAGAGCTCGCGCAGCGCGTGGTGGGCGAGTTC 1725
 576 GlnAlaValAsnAlaValSerGluAlaLeuArgSerArgAlaGlyLeuGlyArgAla 595
 1726 GAAGCGGTGATCGGTCTCGACCGGTCCGCTGCTCGCGCGGCTCGACGAGCCG 1785
 596 GlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeu 615
 1786 AACCGTCCGATGGGTCTGATGTTTATAGCCGCCACCGCGGTGGCAACGAGCTG 1845
 616 AlaLysAlaLeuAlaGluLeuPheAspGluLeuLeuLeuValArgIleAspMet 635
 1846 ACCAAGGCGTCCCGGATGATCTGTTCACGACGAGACCGCGCGGTGGTCCGATCGCATG 1905
 636 SerGluTyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProGlyTyrVal 655
 1906 TCGAGATTTCATGAGAACCACTCCGTGGCCCGGTGATCGCGCACCTCCCGGTATGTC 1965
 656 GlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgArgProTyrCysValIle 675
 1966 GCGTACGACGAGCGGTGCTGTCACCAAGCGGTGCGCGCGCGCCCTACCGAGTGCATC 2025
 676 LeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeu 695
 2026 CTGTTCGACGAGATCGAGAGCGCATCCGACGTGTTCAACGTGCTGTCGAGGTGCTC 2085
 696 AspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIle 715
 2086 GATCAGCGCCGCTGACCGATGTTGTCAGCGCGCACCGTGGACTTCCGCAACACGCTGATC 2145
 716 IleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThr 735
 2146 GTGATGACCTCGAACCTCGGTCTCGAGTATCTGTCAACACGACCGCGCGGCGAGCACCC 2205
 736 MetGluValAlaAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu 755
 2206 GCGCGGTG---CCGAGCAGGTGATCGGATGTTGGCGGCGCATCTCCGCGCGCAATC 2262
 756 LeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeuArgLys 775
 2263 CTCACCGCGTCGACGAGATCATCTGTTCCACCGCTTGCAGAGACGAGATGGGCGG 2322
 776 ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu 795
 2223 ATCTCGACATCCAGTTTCGCGCGGTCTCCCAAGCTGCTCGAGGACCGCAAGATCGTGCTC 2382
 796 AlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly 815
 2383 GATCTCGATGCTGGCGCGCGCTGCTGCTCGCGAGAGGGCTGGGATCGCGCTTACGGC 2442
 816 AlaArgProIleArgArgTrpMetGluLysValValThrGluLeuSerLysMetVal 835
 2443 GCGCGTCCGCTGAAGCGGTGATCCAGCGAGGTCTCCAGGACCGCTCGCGGAGATGATC 2502
 836 ValArgGluGluLeuAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeu 855
 2503 CTGAAGGCTCTGTCAAGGATGGCACCGATCTCGCGATCTCGCGCGAGCGCGGTGCTG 2562
 856 ValTyr 857
 2563 ACCTTC 2568

JLT 11

US-09-812-350-17.p2n.rnpb
 Application 43449
 Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 43449
 LENGTH: 2586
 TYPE: DNA

ORGANISM: No. US20030233675A1toc punctiforme
 US-10-369-493-43449

Alignment Scores:

Pred. No.: 1,81e-175 Length: 2586
 Score: 2247.00 Matches: 446
 Percent Similarity: 72.26% Conservative: 174
 Best Local Similarity: 51.98% Mismatches: 220
 Query Match: 49.38% Indels: 18
 DB: 12 Gaps: 7

US-09-812-350-17 (1-911) x US-10-369-493-43449 (1-2586)

QY 2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
 Db 13 AATCTTAACCAATTTACAGAAAAGCCTGGGAGCGATGCCCATACCCCGCATATGTT 72
 QY 22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAsp 41
 Db 73 AAACATATCAACAACAGCAAAATTCAGAACGCAACCTGATGAAGCGGTGCTAGAACAA 132
 QY 42 -----ProThrGlyLysPheThrGlnAlaIleSerSerAlaGlyGlyGluAsnAla 58
 Db 133 GATGCTGTAGCACTGGGATTCCTCCAAAGCG-----GGTGTAAACCTC 177
 QY 59 AlaGlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 Db 178 CAAAACTCGCGATCGCACT---CAACAAATTTTTCACGTCAGCCAAAGATATCTGCT 234
 QY 79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAla 98
 Db 235 ACTAGCAGTTCCTGTTTACTTGGGACGCGACTAGATACACTTCTAGACCGGCGACGCTG 294
 QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 Db 295 TATCCGAAGATTTCAAGATGAATATATTTCAATTGAACACTTATTCTGCTGCTTACGCT 354
 QY 119 GluAspSerGlnIle---ArgAspLeuLeuAsnGluValValAlaThrAlaArgVal 137
 Db 355 AAAGATGACCGCTTTGGCAAGCTTTATCCAGAAATTCGTTTAGACGAGGCAACTA 414
 QY 138 LysSerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGly 157
 Db 415 AAAGATATTTATTAACAGATTTCCGGGAGCCAA-----AAAGTGACCGACCAAAATCCA 468
 QY 158 AspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla----- 175
 Db 469 GAAGCAATACGAGGACCTCGAAATAATACGGCGGTGACCTTCACAGAAAGCGCGCGTAAA 528
 QY 176 GlyLysLeuAspProValIleGlyArgAspGluLeuIleArgArgValValArgIleLeu 195
 Db 529 GGTCAACTCGATCCAGTGATTTGGCGGGATGATGAGATTCGCCCGACTGTCCAAATTCG 588
 QY 196 SerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
 Db 589 TCTCGCGCACCAAGAAATATCTGTGCTAATTTGGTGAACCGGTGTTGTTAAACTGCT 648
 QY 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
 Db 649 ATCGCTGAAGGATTAGCACGCGGATTATTATTCAGGTGATGATACCCAGTCCCTCAAGAC 708

236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGlu 255
709 CGCAGCTAAATTTCTTAGATATGCTGTTGATGCGGGGCAAAATTCGGGGTAA 768
256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValle 275
769 TTTGAGAACCCCTGAAAGCAGTATTAAGAAGTTACTGAATCTGGCGGCAATATTGTT 828
276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluLysSerMetAsp 295
829 TTATTATTGATGAATTCACACGTTGTTGGCGCTGGTGCACCCCAAGCGCGATGAT 888
296 AlaAlaLeuLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
889 CGCGGTAACTTTGTAACACCGATGTTGGCGGGGTGAATTCGCTGTATTGGGGCGACA 948
316 ThrLeuGluGluThrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
949 ACTCTAGATGAATACCGCAACATATCGAAAGAGTCCGCACCTAGAAAGACGCTCCAG 1008
336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
1009 CAGGTTTATCTCGATCAACCTAGTGTAGAAGATATTTCTGATTTTGGCGGGTGA 1068
356 GluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
1069 GAACGTTATGAACACCCACCGGGGTAAATTTCTGATAGTCTTTGGTCGCGCGCC 1128
376 GlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
1129 GTATTGTGAGTCGATATATTAGCGATCGTCTTACTCTGATAAGCCATTGCTTGGTA 1188
396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
1189 GAGGAACCCCGCCAGATTAATAAGATGAGATCACCTCCAAACAGAAAGAACTCGACGA 1248
416 LeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAsp 435
1249 ATTGATCGCAGATTCTGCAATTTGGAATGGAAGAGCTATCGCTGCAAAAGAAAGCGAT 1308
436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLys 455
1309 GCGGCTTCTCGTGAACGCTAGAAAGACTAGAAAAGAAATTCGCCGATCTCAAGAGAA 1368
456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
1369 CAAGAAACCCCTAAATCTCAATGCAATCTGAAAGAGATATCAATTGACAAATTTCAATCC 1428
476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAsp 495
1429 GTTAAAGAGAGATTGAACGGGTCAATTTAGAGATTCAGCAACGACAGACGATTAAGAC 1488
496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIle-----GlnGluValGlu 511
1489 CTTAACCGAGCTCGGAGTTGAAATACGGTAATTTAACTAGTTTGCATGCTCAGTTGGA 1548
512 SerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsn 531
1549 GCAGTAGAGCTGAATTTGCAAGTGCACCCAAAGAGTGGAAATCACTATACGGGGAAGAA 1608
532 ValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProValThrArg 551
1609 GTCACAGAGCTGATATCTGTAATATTCTAATGACAGGAATTCCTCAATCAGCAG 1668
552 LeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgVal 571
1669 CTGGTGAATCTGAGAAAGAGAACTACTGCAATTTAGAAGATCAACTACGCCACCGTGTG 1728
572 ValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGly 591
1729 ATTGGACAGAGAGAGCAGTCACAGCGGTAGCGATTCAGCAATTCAGCGATCGCGCTGA 1788
592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611

1789 CTGGCGGATCCCAATCGTCCATCGTAGCTTTATTTCTTGGGCTACCGGTGTGGT 1848
612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspGluAsnLeuVal 631
1849 AAAACCGAGTTGCGGAAAGCGTGGCGGCTATATGTTTCGATAGCAAGATCGGTGGTG 1908
632 ArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaPro 651
1909 CGAATCGATATGTCGATATATGAGAAACACCGCGTCTCCGTTTAAATCGGTGGCT 1968
652 ProGlyTyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgPro 671
1969 CCAGGATATGTCGTTTACGAAGCGGCAACATAACAGAGCGATTCCGCCCGCTCT 2028
672 TyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
2029 TACTCAGTATTTCTTTTACGAAATCGAAAGACACACCTCGATGTTTAAATATCTTC 2088
692 LeuGlnValLeuAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArg 711
2089 TTGCAATTTCTGATGATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2148
712 AsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThr 731
2149 AATGCTATTTATTTATCATGACTAGCAACATCGTTCGCAATACATTTCTTTGATGTCGCTGG 2208
732 GlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPhe 751
2209 GATAATGCTACTAGCAAGAAATGCGCTCGAGTCATGGAAGCGATTCGGAATAGCTTC 2268
752 ArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAsp 771
2269 CGTCCAGAAATCTCTCAACCGGATTGACGAATCATCATCTTCCCGTTTATAGATAGAG 2328
772 GlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArg 791
2329 GAATTTGCGCGAGATTGCTGCTGTCAGATCAAAAGATTATCCCAAGATTGGCCGATCGC 2388
792 GlyAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAsp 811
2389 AAATATCCCTCAAGCTCTCAGATGCTGCACTTTTACGAGAGTAGGATATGAC 2448
812 ProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeu 831
2449 CCAGTTTATGGCGCGCTCCACTGAACGGGCGATTTCAGCGAGAGCTAGAACTCAAAAT 2508
832 SerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
2509 GCCAAGCCATCTTGGCGGGTGAATTCACCGATGCGCAACACCATCTTTGTAGAT 2562

RESULT 12

US-10-398-221-5/c
; Sequence 5, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 US
; CURRENT APPLICATION NUMBER: US/10/398, 221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 43980
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:

[illegible]

```

651 ProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArg 670
9161 CCTCCAGGTGTAIGTTGGATATCAGGAAGCGGACACTTACCGAAGCTGTGAAGCGAAC 9102
671 ProTyrCysValLeuLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThr 690
9101 CCATATTCGATGCTTACTTATGATGAATCGAAATCGAAAGCCCATCCAGACGCTATTTAATC 9042
691 LeuLeuGlnValLeuAspAspGlyValGlnLeuThrAspGlyGlnGlyArgThrValAspPhe 710
9041 CTTTACAGTACTTCAGATGTCGGATACGATTCACAAAGCGCGCTTAATGACTTT 8982
711 ArgAsnSerValLeuLeuMetSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeu 730
8981 AAAAACACCGTAATCATATGACGCTAATATCGCTCTAATTTATTTACTTGAAGAAC 8922
731 ---ThrGlyLysValThrMetGluValAlaAlaArgAspCysValMetArgGluValArgLys 749
8921 GAAGAAGGTGAATCTCGCTGGAATAGAAATCAGAC---GTAATGCAAAATTTTGCRAAGT 8965
750 HisPheArgProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSer 769
8964 GAATTTAAACCAAGATCTTAAACCGGTAGATGATTAATTCATTTAAACCACTTACA 8905
770 HisAspGlnLeuArgLysValAlaAlaArgLeuGlnMetLysAspValAlaValArgLeuAla 789
8804 CTCGCTGATATTAAAGGCAATGTGCAAAATTTAGTAGAAGAACTTCAAAATCGCTTGT 8745
790 GluArgGlyValAlaLeuAlaValThrAspAlaLeuAspTyrLeuLeuAlaGluSer 809
8744 GATCAGAAATTTACGATTAATTTTATTCAGATTAACGCAAAACCCCTTTATTCAGAAAGCT 8685
810 TyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThr 829
8684 TATGACCAAGTATACGTCGCGCTCCACTGAACGATATATTTGTCGTCATGCGAAGC 8625
830 GluLeuSerLysMetValValArgGluGluLeuAspGluAsnSerThrValTyrLeuAsp 849
8624 CCTCTTCAAGAAATTTGATCTGTTAAATATTCGCGCATTCCTCCGTTGAATTTGAT 8565
850 AlaGlyAlaGlyAspLeuValTyrArgValGluSerGlyLeuValAspAlaSerThr 869
8564 TTACAGATTAAGAAATTTACTTTTAAAGT-----ACA 8532
870 GlyLysLysSerAspValLeuLeuHisLeuAlaAsnGlyPro----- 883
8531 GAATAACATGGAATACTTTGATTCACGCTCGAAGGGAACCTTTTGTGATTAGTC 8472
884 -----LysArgSerAspAlaAlaGlnAlaValLysLysMetArgLeuGluLeuGlu 901
8471 GTTTAGAAATAAGCGAAATATGGAAGG---ATGGAAGATGAATATGATTAATTCGT 8415
902 AspAspAspAsnGlu 906
8414 CATGCTCAAAGTCAA 8400

```

SULT 13

US-09-812-350-17.p2n.rnpb

Sequence 2058, Application US/10398221

Publication No. US20040018514A1

GENERAL INFORMATION:

APPLICANT: KUNST, Frederik

TITLE OF INVENTION: Listeria innocua, genome and applications

FILE REFERENCE: 344 702 - US

CURRENT APPLICATION NUMBER: US/10/398,221

PRIORITY FILING DATE: 2003-03-27

PRIORITY FILING DATE: 2001-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

PRIORITY FILING DATE: 2000-10-04

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2058

; LENGTH: 3011208

; TYPE: DNA

; ORGANISM: Listeria innocua

US-10-398-221-2058

Alignment Scores:

```

Pred. No.: 1.23e-171 Length: 3011208
Score: 2245.50 Matches: 460
Percent Similarity: 70.59% Conservatve: 193
Best Local Similarity: 49.73% Mismatches: 237
Query Match: 49.35% Indels: 35
DB: 12 Gaps: 12

```

US-09-812-350-17 (1-911) x US-10-398-221-2058 (1-3011208)

```

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 2343102 ATGGATTACAAAATTTTACCACACAGTCCAGCAACATTTGCGGACGCAAAATCTA 2343043
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
Db 2343042 GCTATTGCATCAGAACATCAAGAAATTCAGCTTATACATGTTTAAAGTGTATTAACT 2342983
QY 41 AspProThrGlyLysPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Db 2342982 GAG---AGTGACTTTGCAAAACGCGCATATGATGTAGCA-----GAAGTTAATGTTGAG 2342932
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuPro-----SerGlnSerPro 78
Db 2342931 GCACCTTCAAAAGTAGTAGACGAATCTTTAAGAAAATTCCTGTTGTTTCAGGAAGCGGA 2342872
QY 79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAla 98
Db 2342871 GTGAATATGCGCAAGCAATGAGTCAAGCACTTTTCAACATTAATGAGATGCGAGAAAA 2342812
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
Db 2342811 GAACAAAACCAATTAGACGATGATTTGTATCGACAGAACATCTTTATCTTAGCTGTAATG 2342752
QY 119 Glu-----AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArg 136
Db 2342751 GACCAAAAGTCAAAACCCCATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2342692
QY 137 ValLysSerGluValGluLysLeuArgGlyLysGlyLysValLysValGluSerAlaSer 156
Db 2342691 ATTACGAAGCAATTTTAAATAATCAGAGGA-----GGGAAAAAAGTGACTTCTCAAAAT 2342638
QY 157 GlyAspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGln----- 174
Db 2342637 GCAGAGGAAACTATGAGCTTTAACAATAATGACGAGATTTAGTCGCGGAAGTAAGA 2342578
QY 175 AlaGlyLysLeuAspProValIleGlyArgAspGluGluLeuIleArgArgValValArgIle 194
Db 2342577 AGCGGAAACTTGAATCCAGTGAATGACGCGACGCAAGAAATTCGTAATGTCATCCGAAT 2342518
QY 195 LeuSerArgArgThrLysAsnAsnProValLeuLeuIleGlyProGlyValGlyLysThr 214
Db 2342517 TTTCAAGAAAAACAAATAATATCCAGTTCTAATTTGGTGAACCAAGTGTGTGTAATAACA 2342458
QY 215 AlaValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThr 234
Db 2342457 GCATTTGAGAGGTTTAGCACACGATTTGTGCAAGAAAGCGTTCAGAAAGGATTTGAAA 2342398
QY 235 AspValArgLeuLeuSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGly 254
Db 2342397 GATAAAACAAATTTTCCCTTGATATTGTTGTTCCCTTATTTGTTGGGCTTAAGTATCTGCT 2342338
QY 255 GluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysVal 274
Db 2342337 GAATTTGAGAACGTTTAAAGCAGTACTCCAAAGAAAGTAAACAAAGCGATGACAAAT 2342278

```

275 IleuPheIleAspGluIleHisLeuValLeuGluYalaGlyLysThrGluSerMet 294
2342277 TTGCTCTTTATAGATGAATTCATACAAATGTGCGGCGAGTAAACAGATGAGCTATG 2342218
295 AspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyClnLeuArgCysIleGlyAla 314
2342217 GATGCTGGAAACATGTTGAACCTATGCTTGCAGAGGTGAATTCACCTGATTCGGCGCA 2342158
315 ThrThrIleuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPhe 334
2342157 ACAACCTTTCAGCAATACCGCAATATATCGAAAGAGATGCTGCACTAGAAAGACGTTTC 2342098
335 GlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeu 354
2342097 CAATAAGTACTGTTCCAGAACCAACCGGGAAGATACAGATTTCCATTTTACGTGCTTTA 2342038
355 LysGluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAla 374
2342037 AAAGAACGTTTGAATCATCATGGGTAAATATTCATGATAATGCCTTAGTTCGGCA 2341978
375 AlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeu 394
2341977 GGTAGCTTTCTAATCGTTTACATTCAGGATCGTTTTCACGGATAAAGCAATTCATTTA 2341918
395 ValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAsp 414
2341917 GTAGATGAAGCGGTGCAACTATTCGTGCGAAATTTGATCTATGCCAAGTGAATTCAT 2341858
415 AsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLys 434
2341857 GAAGTAAACAAGAAAGTATGTCAGCTAGAAATTTGAAGAGCGGCAATTAAAGAGAAAGAA 2341798
435 AspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLysLeuArgAsp 454
2341797 GACCAGCAAGTGAACGGCGCTTACAGATGTTACAGCAGAACTAGCTGCTACTATAAGAA 2341738
455 LysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArg 474
2341737 GAAGCAATAAATGAAGTCTTAATGGGAATCCGAAATCCGAAATCCGTAATTCGT 2341678
475 ArgLeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyr 494
2341677 GAAGTCTGTGAACAAATCGATCATCTGGTTCATGATTAAGAGAGAGAGAGCAACTAC 2341618
495 AspLeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIle 514
2341617 GATTAAATAAAGCTGCTGAATTCGCGCAGCTGTAATTCAGCAGTAGAAGAAAGAAATTA 2341558
515 AlaGlnLeuGluGlyThrSerSerGlu-----GluAsnValMetLeuThrGlu 530
2341557 TTACATTTAGAAACGGAATTCGCGAAACAAACAGCCCAAGAGATCGAATTTTACAAGAA 2341498
531 AsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProValThr 550
2341497 GAAGTAAACAGAAATGAATCGCTGAAATTTGTCAGATGCTGCAATACCACTGACT 2341438
551 ArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArg 570
2341437 AAGTTAGTAGAAGAGAGAACCGGAAATCTGTTAAATAGCCGATGTTCTTCATCAAAA 2341378
571 ValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAla 590
2341377 GTAAATGCTCAAGACGCGGTTCAGTTAGTAGTATGATGATGATGATGATGATGATGAT 2341318
591 GlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyVal 610
2341317 GGGATTAAAGATCCAAACAGCAATCCGATCTTTATTTCTTAGGCCCAACCGGTGTT 2341258
611 GlyLysThrGluLeuAlaLysAlaLeuAlaGluClnLeuPheAspAspGluAsnLeuLeu 630
2341257 GGTAAACGGAATAGCAAGCAATAGCCATAATATGTTGATTTCTGAAGATCATATG 2341198
631 ValArgIleAspMetSerGluTyrMetGlnHisSerValSerArgLeuIleGlyAla 650

2341197 ATTGCAATGTATGTCTGAATATACATGAGAAACATTCGCTATCAAGACTTGTGCGGCT 2341138
651 ProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgArg 670
2341137 CCTCAGGTTATGTGATATAGAGAGCGGCAACACTTACCGAAGCTGTAAAGACGCAAC 2341078
671 ProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThr 690
2341077 CATATTGATGTTGCTTACTTGTGATGAATCGAAAGCCATCCAGACGCTATTATATATC 2341018
691 LeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPhe 710
2341017 CTTTACAGTACTTACGATGTCGATTCAGGATTCACAGGCGCTTAATATGACTTT 2340958
711 ArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeu 730
2340957 AAAAACACCGCTAATCATATGACGCTAATATCGGCTCTAATTTACTTTGAAGAAC 2340898
731 ---ThrGlyLysValThrMetGluValAlaAlaArgAspCysValMetArgGluValArgLys 749
2340897 GAAGAGGTGAATCTCGCTGAACTAGAACTAGAC---GTAATGCAAAATTTGCAAGT 2340841
750 HisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSer 769
2340840 GAATTTAAACACGAAATCTTAACCGGTAGATGATATTTCTATTAAACCACTTACA 2340781
770 HisAspGlnLeuArgLysValAlaAlaArgLeuGlnMetLysAspValAlaValArgLeuAla 789
2340780 CTCGCTGATATTAAGCAATTTGCGAAATTTAGTAGAAGAACTTCAAATTCGCTTGT 2340721
790 GluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSer 809
2340720 GATCAGAAATTAAGTAATCTATTTCAGATAACGCAAAAGCCCTTTATTCAGAAAGAGCT 2340661
810 TyrAspProValTyrGlyAlaArgProIleArgTyrMetGluLysLysValValThr 829
2340660 TATGACCCAGTATACGTCGCGCTCCCTCAGAAACGATATATTGTGCGTCATGTGCAACG 2340601
830 GluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
2340600 CTTCTTTCGAGAGAAATTTATCTGTAATAATATGCGCAATTCCTCCGTTGAAATTTGAT 2340541
850 AlaGlyAlaGlyAspLeuValTyrArgValGluSerGlyGlyLeuValAspAlaSerThr 869
2340540 TTACAAGATAAGAAATTTACTTTTAAAGTC-----ACA 2340508
870 GlyLysLysSerAspValLeuIleHisIleAlaAsnGlyPro----- 883
2340507 GAATAACATGGAATGTTTGTATTCAGCTCGAAGGGAACCTTTTGTGATGATGTC 2340448
884 -----LysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluLeuGlu 901
2340447 GTTTGAAATTAAGCGAAATGGAAG---ATGAGGAAGATGAAATAGTATTATTAATCGT 2340391
902 AspAspAspAsnGlu 906
2340390 CATGTCTCAAGTGAA 2340376
RESULT 14
US-10-369-493-46242
; Sequence 46242, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10 (52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493

Db 784 ATCAAATCGTTTTAAAGAAGTGTGAAGAAAGTGAACCTCCCAATATTTATTTATTCGTAGAT 843

QY 280 GluIleHisLeuValLeuGlyAlaGlyLys---ThrGluGlySerMetAspAlaAlaAsn 298

Db 844 GAAATGCATCTTTAATGGCGCAGAGTTCGGTGTGTGAAGGTGGCATGGACGCTGCCAAT 903

QY 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuAryCysIleGlyAlaThrThrLeuGlu 318

Db 904 TTGCTGAAGCCAATGCTTGCACGTGTAAGCTTCACTGTATTGGTGCTACTACTTTGGCT 963

QY 319 GluTyrArgLysTyValGluLysAspAlaAlaPheGluAryArgPheGlnValTyr 338

Db 964 GAATACAGAANAATCATCGAAAGAATGCTGCTTTCGAACGTAGATTCCAAATCATCTTG 1023

QY 339 ValAlaGluProSerValProAspThrIleSerIleLeuAryGlyLeuLysLeuLysTyr 358

Db 1024 GTAAAAGAGCCTTCTATTGAAGATACGATTTCTATCTTCGTGCTTTAAGAAAGATAT 1083

QY 359 GluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378

Db 1084 GAGGTTCAATCATGGTGTTTACTATCTCTGACCGTGCCTTGGTTACTGTGCTGCCCATCTCGT 1143

QY 379 AlaArgTyrlleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398

Db 1144 TCCGGGTACCTAACTCTCTCGTCGTCTTCAGATTTCTGCCATCGACTGGTTCGATGAAGCT 1203

QY 399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418

Db 1204 CGCGCTCGAGTTCGTGTACTCTCGTGAATCTCAGCCGCAAGTTCCTTGATAACCTTGAGAGA 1263

QY 419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluAryGluLysAspLysAlaSer 438

Db 1264 AAGCTTCGACAACTCCCGGTGAAATTCGTGCTTTAGAGCTGTGAAGACGAAAGCTTCC 1323

QY 439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgLysLeuGlnPro 458

Db 1324 AAAGAACGTCTCAAGGCTGCTGCTAAAGAGCTGAACAAGTGGAGNAGAAACAAGACCT 1383

QY 459 LeuThrMetLysTyArgLysGlyLysGluAryGleAspGluIleArgLysGln 478

Db 1384 ATTCGTGAGAAGTATGAATTAGAAAATCTCTGTGGAAGTGAACCTCAAGATGCTAAGCGT 1443

QY 479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluAryArgTyrAspLeuAlaArg 498

Db 1444 CGCTTGATGAGTTGAAGCTAAGCTGAGATGAGCGGCGAGCTGTAATGATTTCACTCTT 1503

QY 499 AlaAlaAspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518

Db 1504 GCTGCTGATTGAATACTATGTTATCTCTGATTTGCAAAAACGATTGTAATATTAGAG 1563

QY 519 GlyThrSerSerGluGluAsnVal----- 526

Db 1564 CACAAAAAGAAAGCTGATGCAAGCGATGTGCAACCGCTCAACCGGCTCAGAACCT 1623

QY 527 MetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGly 546

Db 1624 TTACTTTATGACGTAGTGTGGGCCGATCAATCAACGAATCTGTTCAGAGTGGACTGGT 1683

QY 547 IleProValThrArgLeuGlyGlnAsnGluLysGluAryGleLeuIleGlyLeuAlaAspArg 566

Db 1684 ATTCCGGTTTACTCGTTGGAAGACACTGAAAAAGGAACGACTCCTTAAACATGAGAGGTT 1743

QY 567 LeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeu 586

Db 1744 CTCAGTAAACAAGTAATGGTCAAAATGAGCAGGTTACTGTCAGTTGCAACGCCAATTCGC 1803

QY 587 ArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGly 606

Db 1804 TTATCTCGTGGTCTTCTGATCCTTAATCAACCAATCGCATCTTCTTCTTTGTCGGT 1863

QY 607 ProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnPheAspAsp 626

2.06e-174	Length:	2715
2234.50	Matches:	460
66.52%	Conservative:	158
49.52%	Mismatches:	258
49.11%	Indels:	53
12	Gaps:	11

2-350-17 (1-911) x US-10-369-493-46242 (1-2715)

6	PheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAlaValAsnAlaGly	25
16	TTTACTGACAAAGCGCAAGACATTTGCTGATCGGTACTCAATTGCAACATTTATGGT	75
26	HisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSerAspPro	42
76	CATTCAAAATTAACCCCTATTCAATTGCTGCTCTCTCTTTGTCGACAGTGCAGTAAAC	135
43	--ThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGlnSer	61
136	GGTACTACTTACTTCGCGCACCATCGTCGCAAAAGCTGGTGGTAT	189
62	AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp	81
190	TTTGAGCGAAGTGTTACTAGTCGGTTAGTTCGCTGCTGCTCAGGATCCACCTCCAGAA	249
82	AspIleProAlaSerSerSerLeuIleValValIleArgArgAlaGlnAlaGlnLys	101
250	CAAGTGACTCTGTCACCTGAAAGTGGCAAACTGCTCGAAATGCTCATGAATCTCAAAA	309
102	SerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSer	121
310	ACACAAAAGATTCCTTACATTCGTCAGATCATTTTATCGCTGTGTTCTAAGACGAC	369
122	GlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArgValLysSerGluVal	141
370	ACGCTAAAGAGTTATTGGCAGAACCGGTGTCACCCGAAAGCGTTTGAATTTGCTGTA	429
142	GluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsnPhe	161
430	AATAACCTCGTGGT-----AATAAAGAATCGATTCAAAGNAATCCGGAAGAGATT	483
162	GlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLysLeuAsp	179
484	GATGCTCAATAATAAATTTACTGTGTGATCTTACAGAGCTCGCTAGAAATGGTCAACTGAT	543
180	ProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThr	199
544	CCAGTTATCGACGTCAAGATGAATTCGTCGTACAAATTCGGGTACTTTCTAGAGAAGC	603
200	LysAsnAsnProValLeuIleGlyGluProGlyValGlyIleThrAlaValValGluGly	219
604	AAAAATAATCTCAATTCATTGCTGAGCTGGTGTAGGTAAAGCTCCATTGCTGAAGGA	663
220	LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle	239
664	CTTGCTCGTATATAATTCAGCATGATGTCCCGCTAATCTTTGAAATTTGAAGCTTCTC	723
240	SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArg	259
724	AGCCTGGACGTGGAAAGCTTAGTTCGGGCTCCAAATTCGTGGTGAATTTGAAGAGCGT	783
260	LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp	279

1864 CTTTCGGTACTGCTAAGACTTTATTGACAAAGCATTGTCATCATTTATGTTGATGAT 1923
627 GluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArg 646
1924 GAAATGCAATGATTCGTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1983
647 LeuIleGlyAlaProGlyTyrValGlyHisGluGluGluGluGluGluGluGluGlu 666
1984 CTATTGGAGCACCACCGGTTATGTTGGATGATGATGATGATGATGATGATGATGATGAT 2043
667 ValArgArgProTyrCysValIleLeuPheAspGluValGluGluGluGluGluGlu 686
2044 CTTTCGGTACTGCTAAGACTTTATTGACAAAGCATTGTCATCATTTATGTTGATGAT 2103
687 ValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeuThrAspGlyGlnGlyArg 706
2104 GTATTGACTGATCTTTTGGAGTACTAGATGATGATGATGATGATGATGATGATGATGAT 2163
707 ThrValAspPheArgAsnSerValIleMetThrSerAsnLeuGlyAlaGluHisLeu 726
2164 GTTGTGATGCCAAGATGCTGTTATCATTTATGATCTTAACCTTGGGCGCTGAATCTTG 2223
727 LeuAla-----GlyLeuThrGlyLysValThrMetGluValAlaAlaArgAspCysVal 743
2224 ACAACAGACAATGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
744 MetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluIleVal 763
2281 ATGAATCCGATCGTGGCTTCTTCGTCGCGAATTTTGAATCGTATCTCTCAATCGTT 2340
764 ValPheAspProLeuSerHisAspGlnLeuArgLysValAlaAlaArgLeuGlnMetLysAsp 783
2341 ATCTTTAATCGTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
784 ValAlaValArgLeuAla-----GluArgGlyValAlaLeuAlaValThrAspAlaAla 801
2401 GTTCAGAAACCGCTTCAATCCCAATCATCGCTCAATTAATAATCGAGGTTAGCGCAAGCT 2460
802 LeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArgArg 821
2461 AAAGACCTTCTTGGAGCGCTGTTTATTCCTGCTTATGCTGATGCTGATGCTGATGCTG 2520
822 TrpMetGluLysValValThrGluLeuSerLysMetValValArgGluGluLeuLeuAsp 841
2521 GTCAATCAAAATCAAGTTTGAATCCCAATGCTGCTTAAATTTGAATGCGCACTTGGC 2580
842 GluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSer 861
2581 GATAAGGAGACTGCTCATGCTGCTAGTCCAAATGCAAGATATTGTCAACCAAAAT--- 2637
862 GlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAlaAsn 881
2638 -----CATGAGCTAAT 2649
882 GlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIleGlu 901
2650 GCAAAACGCTCTGCTGAT-----ATTGATGATGATGATGATGATGATGATGATGATGAT 2688
902 AspAspAspAsnGluGluMetIleGlu 910
2689 GACGATGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2715

ULT 15
sequence 44640, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 44640
LENGTH: 2588
TYPE: DNA
ORGANISM: SYNECHOCOCCUS SP. WH 8102
US-10-369-493-44640

Alignment Scores:
Pred. No.: 6,56e-174 Length: 2588
Score: 2228.00 Matches: 455
Percent Similarity: 71.88% Conservative: 166
Best Local Similarity: 52.66% Mismatches: 221
Query Match: 48.97% Indels: 23
DB: 12 Gaps: 10

US-09-812-350-17 (1-911) x US-10-369-493-44640 (1-2588)

QY 1 MetAsnPro-----GluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHis 18
Db 1 ATGCAACCCACAGCCGACAGTTCACCGAACAGCCCTGGCGCCGATCGTCGGCGGCCAG 60
QY 19 GluLeuAlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeu 38
Db 61 CAACTGCCCGGCTCTCAGGCACAGCAGCTGCGAAACCAACATCTGCTGCTGCGGCTG 120
QY 39 IleSerAspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAla 58
Db 121 CTTCCAGAG--AACGGACTGGCGCGGAATCCTCAGCAAAACCGCGCTTGAC----- 171
QY 59 AlaGlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
Db 172 GTCACCACTTCGAGGCCAGTGTGAAGCCATCTCAACGGCTTCCAGCTTGGCTCG 231
QY 79 ProProAspAspIleProAlaSerSerSerLeuIleValIleArgAlaGlnAla 98
Db 232 GCACCCGACTCGGTGTTCTTGGCCGATCCCTCAACAGCCCTCGACGCGCGAGCAG 291
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGlyLeuLeu 118
Db 292 CGCGGGATGATTCGCGACAGCTTCATCGGCATCGAGCATCTGCTTCTGCGCTGGCG 351
QY 119 GluAspSerGlnIle---ArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
Db 352 GAGGATGATCGTGGCGCGCAACTGCTCAGCCAGCGCGGGGTGACCAACACACTC 411
QY 138 LysSerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGly 157
Db 412 AAGGAGCCATCACGCGCGTGGCGCAACCAA-----ACGGTACACAGCAGCAACCT 465
QY 158 AspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla----- 175
Db 466 GAGGCCACTCAGATCCTTGGCGAATACGCGCCGATCTCAGACAGCGCGCGCGCGAT 525
QY 176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgValValArgIleLeu 195
Db 526 GGTCACTGATCGGTGATCGACGGGATGATGATGATGATGATGATGATGATGATGATGAT 585
QY 196 SerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyThrAla 215
Db 586 AGCCCGCGCACCAAGAACACCCGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 645
QY 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
Db 646 ATCGTCGAGGGCTTGGCACAGGATGATGATGATGATGATGATGATGATGATGATGAT 705
QY 236 ValArgLeuIleSerLeuAspMetGlyValAlaLeuValAlaGlyAlaLysTyrArgGlyGlu 255
Db -----

706 CGACAGCTCATCTACCTCGACATGGTGGTCCCTGATGCGCGGGGCGAATAATACCGGGGTGAG 765
256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIle 275
766 TTCGAAGAACGGCTCAAGCGGTGCTGAAGAGGTCAACCACTCCGATGACAGATTGTG 825
276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
826 CTGTTTCATCGATGATCCACACCGGTGTGTGGCGCTGTGGCGCACGGGTGTGCTCATGGAC 885
296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
886 GCCAGCATCTGCTGAACACCGATGCTGCGCGGGGTGAATGCGCTGCATCGGGGGCCACC 945
316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
946 ACCCTTGATGACCGCTGACACATCGMAAAGGATCCCGCCCTGGAGCGACGCTTTCAG 1005
336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
1006 CAGGTGTGGTGGATCAGCCACAGTCCCGGACAGATTTCAATCTCTGCGGGGCTGAAG 1065
356 GluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
1066 GAGGCTACCGAGGTACACCGCGGTACGATTCGCCGACGCGCTTTGTGGCGCGGCC 1125
376 GlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
1126 ATGCTCAGCAGCAGATACATACCGATCGCTTCTCGCGGACAGGCAATCGACCTGGTG 1185
396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
1186 GATGAATCCGCGCGCGGTGAAATCGAATCAGATCACCCTCCAAACCGGAGCAGATCGACGAG 1245
416 LeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAsp 435
1246 ATCGATCGCAAAATCTCGAGCTGGAGATGGAATAACTCTCCCTCGCGCGGTGAGTCCGAC 1305
436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgLys 455
1306 AGTGCCAGCAGGAGCGGTGCAACCGGATGAACCGGAAATTTGGCGCAATCTGGGTGAACAA 1365
456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
1366 CAGAGCAGCTCAATGCCCGAGTGGCGAGCGAAAGGTGCCATTGATCAGCTCTCGGCT 1425
476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrArg 495
1426 CTCAGGAGGAGATCGAACGGGTGCAGCTGCAGTGGAGCGACGCGCAACGCAATTACGAC 1485
496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAla-----IleGln 508
1486 CTCAAAGGAGCAGCTGAATCGAATACGGAATCTCGCCACGCTGCAACGCCAGCTGCAG 1545
509 GluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeu 528
1546 GAGCAGAGACCTCTTGGAGAGCAGGAGCGGAACA-----GACAAAC-CTGCTG 1595
529 ThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIlePro 548
1596 CGGAGGAGGTCAACGAGCAGACATCGTGAAGTGAATGCAAGTGGAGCGCGATCCCC 1655
549 ValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHis 568
1656 GTGCTCGCGCTGTGAGAGCGAAATGGAGAAACTGCTGAGTGGAGGACGACCTGCAT 1715
569 LysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSer 588
1716 CAACGGGTGATCGACAGAACCGAGCGGTCACTGCGGTGGCGGCGGATTCACGCTCA 1775
589 ArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThr 608
1776 AGGCGAGGCTCAGCGATCCCAACCGGCGCCATCGGAGCTTCTCTCTCGGCGCGAG 1835

QY 609 GlyValGlyLysThrCluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsn 628
Db 1836 GGTGTGGCGAAGCAGCACTGTCCAGGCGCTGGGCAATCGCTGTTCGACAGCGACGAC 1895
QY 629 LeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIle 648
Db 1896 GCCATGTGTCGATCGACATGTCGAGTACATGGAGAAACACACCGTGGAGCGCATGATC 1955
QY 649 GlyAlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 668
Db 1956 GGTGCCCTTCGCGGTATGTGGGTATGAAGCGCGCGCCAGCTCACGAGCGGTACGG 2015
QY 669 ArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPhe 688
Db 2016 CAGCGGCTTACCGCGTGTATCTGTTCAGCGAGGTGGAGAAAGCCACCCCGATGTGTTTC 2075
QY 689 AsnThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrVal 708
Db 2076 AATGTGATGCTCGACATCTCGATGACGCGCGGTTCACGCGCGCGCGCGCGCGGTG 2135
QY 709 AspPheArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAla 728
Db 2136 GATTTTCAACCAACACCGTGTGTATCTCACCAGCAACATCGCGCAGCAATCGATTCTTTGAG 2195
QY 729 GlyLeuThrGlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArg 748
Db 2196 ---CTGGTGGGATCCGAGCAACACACGCGGATGGAGCAACGGGTCAACGAGCGGCTC 2252
QY 749 Lys---HisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspPro 767
Db 2253 AAGGCCAAATTCGTCGCCGAATTTCTCAACAGGCTCGACGATCAGATCATCTTCGCGCAGC 2312
QY 768 LeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArg 787
Db 2313 CTTGAAAAGAGGAGCTGCGCGGATCGTGTCACTGAGGTGGAGCGTCTGCGGTGCGGA 2372
QY 788 LeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAla 807
Db 2373 CTGGAGCAACCAAGCTGGACCTTCAGCTCAGCGCAATCGCGCGCATTCGCTGGCCACC 2432
QY 808 GluSerTyrAspProValTyrGlyAlaArgProIleArgTyrMetGluLysLysVal 827
Db 2433 ATTGGCTTTGACCGGTTTACCGAGCAAGACCCCTCAACCGCGCATCCAGCGGAGCTG 2492
QY 828 ValThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyr 847
Db 2493 GAACACCAATCGCCAAAGCGATCTCTGGCGGACAGCTCAGTGGAGGACAGACGCGTGCAG 2552
QY 848 IleAspAlaGly 851
Db 2553 GTTGATGCGAGT 2564

Search completed: February 13, 2004, 10:47:34
Job time : 14451 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - nucleic search, using frame_plus_p2n model

on: February 13, 2004, 01:40:31 ; Search time 3827 Seconds
(without alignments)
5785.570 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

quence: 1 MNPEKTKHTNTATTAHEL.....VKORIEEDDDNEEMIED 911

oring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

urched: 22781392 seqs, 12152238056 residues

al number of hits satisfying chosen parameters: 45562784

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

mand line parameters:

DEL=frame+ p2n.model -DEV=xlp
/cgm2 1/USPTO spoal p/US09812350/runat 10022004 101515 13361/app query.fasta_1.1095
=EST -QFMT=fastap -SURFIX=p2n.rst -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
ITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
CALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
TPT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
ER=US09812350 @CGN 1 1 2874 @runat 10022004 101515 13361 -NCPU=6 -ICPU=3
MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
V TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
APEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

abase :

EST: *
1: em estba: *
2: em esthum: *
3: em estin: *
4: em estmu: *
5: em estov: *
6: em estpl: *
7: em estro: *
8: em htc: *
9: gb est1: *
10: gb est2: *
11: gb htc: *
12: gb est3: *
13: gb est4: *
14: gb est5: *
15: em estfun: *
16: em estom: *
17: em gss hum: *
18: em gss inv: *
19: em gss pln: *
20: em gss vrt: *
21: em gss fun: *
22: em gss man: *
23: em gss mus: *
24: em gss pro: *
25: em gss rod: *
26: em gss phg: *
27: em gss vrl: *
28: gb gssI: *

29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3574.5	78.6	3163	11	AY109444 Zea mays
2	1518.5	33.4	2049	11	AY108805 Zea mays
3	1268.5	27.9	843	14	CB349719 CAB2SG000
4	1261.5	27.7	846	14	CB894504 EST647236
5	1240.5	27.3	837	14	CB894478 EST647270
6	1188.5	26.1	812	14	CB893989 EST646781
7	1188.5	26.1	818	10	BG648761 EST510380
8	1172	25.8	783	10	BE642455 Cri2.6.B1
9	1169.5	25.7	804	14	CB893763 EST646555
10	1152	25.3	740	13	BQ39239 WHE4163.H
11	1115	24.5	722	14	CB610803 ALBDO000
12	1115	24.5	750	28	BH603944 BOGTL30TF
13	1114.5	24.5	727	13	BQ993972 QGF5P11.Y
14	1111	24.4	736	10	BG648385 EST510004
15	1098.5	24.1	729	13	BQ862449 QGC21B24*
16	1098.5	24.1	747	10	BG647349
17	1097	24.1	783	14	CB893376
18	1089	23.9	709	29	BZ963117 PUGFU64TD
19	1070	23.5	686	13	BQ869697 QGD5P04.Y
20	1054	23.2	753	29	BZ963112 PUGFU64TB
21	1044	22.9	863	28	BH465756 BOGTL04TF
22	1001.5	22.0	703	10	BF480849
23	991.5	21.8	818	28	BH250711 BOGAQ17TF
24	989	21.7	1100	29	CNS06844
25	983	21.6	1036	29	CNS06B91
26	969	21.3	900	28	BH163888
27	966.5	21.2	892	28	BH148894
28	964	21.2	872	28	AZ671816
29	957	21.0	895	28	AZ528202
30	956	21.0	617	12	BJ291097
31	953	20.9	689	28	BH423861
32	953	20.9	1798	11	AY103877
33	952.5	20.9	894	28	AZ679340
34	945	20.8	625	9	AW257317
35	935	20.5	1015	28	AF106743
36	933	20.5	903	28	BH132749
37	930	20.4	863	28	AZ679319
38	928	20.4	935	14	CB894408
39	926.5	20.4	813	28	AZ528443
40	926	20.4	598	14	CB894119
41	923.5	20.3	890	28	AZ548311
42	923	20.3	621	13	BH892034
43	918.5	20.2	916	28	BH164525
44	917	20.2	692	10	AW930379
45	915.5	20.1	947	28	BH130876

ALIGNMENTS

RESULT	1	3163 bp	mRNA	linear	HTC 17-OCT-2002
LOCUS	AY109444				
DEFINITION	Zea mays CL1518.1 mRNA sequence.				
ACCESSION	AY109444				
VERSION	AY109444.1	GI:21213165			
KEYWORDS	HTC.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	1	(bases 1 to 3163)			

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

AUTHORS

Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whittitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M., and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 3163)
Coe, E.H.

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES

source

Location/Qualifiers
1..3163
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630374"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"

/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

626 a 846 c 942 g 490 t 259 others

BASE COUNT

TGCIN

Alignment Scores:

ad. No.:

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

1 MetAsnProGluysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20

154 ATCAATCCGACAACTTACCACAGACGAGCGCATCGTNNNNCGACGAGATT 213

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40

214 GCGGTGAGCGCGGCCACGCGCAGCTCACGCGCTGCACCTGCGCGCATGCTGCTGCG 273

41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60

274 GACAAGGGCGGCATCTCGCGCAGGCCATCANNNNNNNNNNNNNNNNNNNNNNNN 333

61 SerAla-----GluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78

334 NNNNNNNCGTTCGAGCGCGTCTGCAACAACTCGCTCAAGAACCTCGCGTCCGCGCCG 393

79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgAlaGlnAla 98

394 CCGCGGATCCGTCGCGCGCTCCACGCGCTGATCAAGGTTCATCCGCGCGCGCAGTCC 453

99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118

454 GCGGAGAAGAAACGGGGACTCGACCTCGCGCTCGACCGCTGCTGCTGCTGCTGCTC 513

119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138

514 GAGGACTCCAGATCTCCGACTGCTCAAGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 573

139 SerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSerGlyAsp 158

Db 574 NNNNGCTTGGANN 633
QY 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
Db 634 ACCAACTTCCAGCGCTCAAGACATACCGCGGACCTCTGTCGAGCAGCGCGGAGCTG 693
QY 179 AspProValIleGlyArgAspGluGluLeuArgValValArgValValArgValArg 198
Db 694 GACCCCGTCATCGCGCGCAGCAGGAGATCCGCGCGTCTGTCGCGCATCTCTCGCGCGC 753
QY 199 ThrLysAsnAspProValLeuLeuGlyGluProGlyValGlyLysThrAlaValGlu 218
Db 754 ACCAAGAACACCCCGCTCTCATCGCGAGCCCGCTTGGCAAGACGCGCTGCTGAG 813
QY 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
Db 814 GCGCTTCGCGCAGCGCATCTGTCGCGGCGACGTGCCAGTAACCTCTCTCGAGCTCGGCTC 873
QY 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 258
Db 874 ATCGCGCTCGACANN 933
QY 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
Db 934 CGGCTCAAGCGCGTCTCAAGAGGTGAGAGGCCGAGGAGGAGGTTCATTTCTTTCATC 993
QY 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAsn 298
Db 994 GACGAGATACACCTCGTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1053
QY 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeuGlu 318
Db 1054 CTGTTCAAGCCATGCTGGCGAGGGGACAGTCAAGTGCATCGCGCCACCCAGCTGGAG 1113
QY 319 GluTyrArgLysTyrValGluLysAspAlaPheGluArgArgPheGlnGlnValTyr 338
Db 1114 GAGTACCGCAAGTACGTGGAGAGGACGCGCTTCAGCGCGCGGTTCCAGCAGGTGTTTC 1173
QY 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
Db 1174 GTCGCGAGCGCGAGCGTGCACACCGTCAGCATCTGAGCGGACTCAAGGAGAGTAC 1233
QY 359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuLeuLeuAlaGlnLeuSer 378
Db 1234 GAGGCGCACCATGCGGTGAGATCCAGACCGCGCTCTGTCGCGCGACAGCTCTCC 1293
QY 379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
Db 1294 GCGAGGTACATCATGCGTCCGACCTGCCTGACAAAGCCATACCTGTCGAGCAGGCC 1353
QY 399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluLeuAspLeuGluArg 418
Db 1354 TCGCGCCATGTAGGGTGCAGCTCGACAGCAGCGAGGAGATTGATTAACCTGGAGAGG 1413
QY 419 LysArgMetGlnLeuGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
Db 1414 AAGGATCCAGTTGAGTTGAGTTCACCGCTCGAGAGGAGGAGGAGGAGGAGGAGG 1473
QY 439 LysAlaArgLeuLeuValArgLysGluLysGluLeuAspLeuArgAspLysLeuGlnPro 458
Db 1474 AAAGCCCGCTGATTCAGGTTCAGAGGAATTCGACATCTGAGGAGCAGCTGAGGCC 1533
QY 459 LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluLeuArgGluLeuGln 478
Db 1534 CTGACCATGAAGTANN 1593
QY 479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArg 498
Db 1594 CGCGCGAGGAGTTCAGTTCCACCTGCGAGGAGCGCGCGGATGCGGATGCGGCTG 1653
QY 499 AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518

```

1654 GTGGCGGACCTCAAGTACGGCCCTCCAGGAATCGACGGCGCGATCTCCAAGCTN--- 1710
519 GlyThrSerSerGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAla 538
1711 NNNNNNNNNNNNNNNNNNTGATGCTACCGAAACCGTCGGCCCTGAACAATTGCA 1770
539 GluValValSerArgTyrThrChlyleProValThrArgLeuGlyGlnAsnGluysGlu 558
1771 GAGGTGGTGAGCGGTGGACGGGTATTCCAGTGACCGCGCTTGGCGAAGCAACAAGGAG 1830
559 ArgLeuIleGlyLeuAlaAspArgLeuHisIleArgValValGlyGlnAsnGluAlaVal 578
1831 AGGCTGGTGGCTGGCTGGCTGACAGGCTTACAGAGGGTGGTGGCCACAGAGGCTGTG 1890
579 AsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnPro 598
1891 AGCGCGGTGGACGAGCGGTGCTGAGGTGGAGGGCGGTCTTGGCAGGCCACAACAGCCC 1950
599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyGlyThrGluLeuAlaIleAla 618
1951 ACTGGCTCGTTCTCTCTGGGTCCGACTGGCGTGGGGAACCTGAGCTGGCCAGGCC 2010
619 LeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyr 638
2011 CTAGCCGAACAGCTGTTCCAGCAGAGAACCTTCTGTCCGATCGACATGTCGAGTAC 2070
639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGlu 658
2071 ATGAGCAGCAGCTCGGTGCTCGCTCATCGAGCACCACCTGGCTACGTCGGCATGAA 2130
659 GluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAsp 678
2131 GAGGTGGCGAGCTGACTGAACAAGTGAAGGAGGCGGTACACGCGTATCTGTTCCGAC 2190
679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly 698
2191 GAGTCTGAGAGGCCCATGTCGCGGTGTTCAACACCTGCTCCAGGTCCTCGACAGCGC 2250
699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMetThr 718
2251 AGGTGACGATGGCGAAGCGGAGCGGAGCTGACCTCAGGAACACCGTATCATGACA 2310
719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
2311 TCGAAGCTCGGCGCGGAGCAGCTCTCGTGGAGTGGTGGCGAAGAACTCCATGAAGGTC 2370
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2371 GCTCGCATCTGGTCAATGACGAGGTGAGGAGGCACATTCGCCCTGAGCTGCTGAACCGT 2430
759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2431 CTGACAGAGATCGTGATCTTCGATCTCTCTCCACAGCAGCAGCTGAGGAAGGTGCTCGC 2490
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2491 CTTGATGAGAGATGGCGCGTCCGCTTTCGGAAGGGGATCGCTCGGTGTGACC 2550
799 AspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818
2551 GAGCGCCATTTGACATCATCTGTCTCTCTTACGATCCGTTGACGGCGCGGCCA 2610
819 IleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
2611 ATCAGGAGGTGGATCGAAGAAGAGGGTGGTGGCAGCGTGTGGAAGATGCTGATCCAGGAG 2670
839 GluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuVal 856
2671 GAGTTCAGCAGAACTCCAGGTTTACATCGACCGCGCGCCCGGCAAGCAGCGTGGTC 2730
857 TyrArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
2731 TACAGGTGGACCGGCGCGTCTGTTGTAACGCTGAGACGCGGGATGAGTGGGACATC 2790

```

```

876 LeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMet 895
2791 CTGATCAGGTCCCAACAGCTCCACGAGGAGCGCTGCGCAGCGCGTCAAGAAGATG 2850

896 ArgIle---GluGluIleGluAspAspAsnGluGlu 907
2851 AGGATCATGAGGAGGAGGAGGAGGAGCGCATGCGACGAGGAG 2889

RESULT 2
AY108805 Zea mays PC0077883 mRNA sequence. linear HTC 17-OCT-2002
LOCUS Zea mays PC0077883 mRNA sequence.
ACCESSION AY108805
VERSION AY108805.1 GI:21212055
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2049)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2049)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
FEATURES
Location/Qualifiers
1..2049
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:634307"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT 615 a 409 c 565 g 460 t
ORIGIN
Alignment Scores:
Pred. No.: 1,92e-134 Length: 2049
Score: 1518.50 Matches: 318
Percent Similarity: 65.99% Conservative: 138
Best Local Similarity: 46.02% Mismatches: 166
Query Match: 33.37% Indels: 70
DB: 11 Gaps: 8
US-09-812-350-17 (1-911) x AY108805 (1-2049)
876 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGluGly 219
2851 AAGAACATCCCTGCTTATTTGGAGAGCGCTGGTGTGGAAGAACTGCAATTGCTGAAGG 73
2851 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 239
74 CTGTGCTAGCGCATTTCTTACAGGCGATGTACCTGAAACAATAGAGGAAAGGTCATA 133

```

LOCUS	CB349719	843 bp	mRNA	linear	EST 10-APR-2005
DEFINITION	CB2SG0007 IIIAf All Cabernet Sauvignon Berry - CAB2SG Vitis				
ACCESSION	vitis cDNA clone CAB2SG0007 IIIAf All 5', mRNA sequence.				
VERSION	CB349719				
KEYWORDS	CB349719.1 GI:28970686				
SOURCE	EST.				
ORGANISM	Vitis vinifera				
	Vitis vinifera				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.				

1 (bases 1 to 843)
 Goes da Silva,F., Iandolo,A., Lim,H., Baek,J., Jones,K. and Cook
 D.
 Expressed sequence tags from cabernet sauvignon berries at various
 developmental stages
 Unpublished
 Contact: Douglas Cook, PhD
 CAES Genome Facility
 UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561
 Fax: 530 754 6617
 Email: drcookeucdavis.edu
 Seq primer: ACGGTACCGACATATGCC.
 Location/Qualifiers
 1. .843
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultivar="Cabernet Sauvignon"
 /db_xref="taxon:29760"
 /clone="CAB2SG0007.IIIaF.All"
 /sex="Hermaphrodite"
 /dev_stage="Veraison"
 /lab_host="DHSalpa"
 /clone_lib="Cabernet Sauvignon Berry - CAB2SG"
 /note="Organ: Berry; Vector: pDNR; Site 1: Sfil; Site 2:
 Sfil; CAB2SG is a cDNA library of Cabernet Sauvignon Clone
 8 berries. Samples were collected at veraison (transition
 between stage II and stage III of berry growth) from
 field-grown vines 60 days after full bloom. Berries were
 still green but soft. Sampled vines were located at the
 University of California, Davis, experimental vineyard.
 cDNAs were made by oligo-dT priming and directionally
 cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGCTATCAACGAGAGTGCCTTACGCCGG-3' and
 5'-ATCTAGAGCGCGAGCGCGCATG-dT(30)NN-3'. Library was
 constructed using the Clontech Creator SMART kit and
 size-selected to contain the 0.5-3 kb size fraction."

FEATURES
 source

E COUNT 218 a 166 c 251 g 207 t
 GIN

gment Scores:

d. NO.: 5,238-111 Length: 843
 re: 1268.50 Matches: 250
 cent Similarity: 95.73% Conservative: 19
 t Local Similarity: 88.97% Mismatches: 11
 ry Match: 27.88% Indels: 1
 14 Gaps: 1

09-812-350-17 (1-911) x CB349719 (1-843)

498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 2 AGAGCTGCTGATTGCGATATGGCGCAATTCAAGAGTGGAGGCTGTATAGCAACCTT 61
 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 62 GAAGCAACC---ACTGATGAGAACAATGATGTTAAACGGCACTGTTGGGCTGAGCAGATT 118
 538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlnAsnGluIys 557
 119 GCTGAGGTGGTGGAGCGCTGGACTGGCATCCTGTCTAGCTTGGACAAATGATATA 178
 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisIysArgValValGlyGlnAsnGlnAla 577
 179 GAAAGTTTAAATTGGGCTTGGAGAGGCTGCATCAGAGTGGTGGACAGGATCAGGCA 238
 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 239 GTCACTGCTGTGGCAGAGCGCTCTTTGATAGTCGAGGCTGGGTTAGGAAGGCCAACAA 298
 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyIysThrGluLeuAlaIys 617

Db 299 CCCACTGGTTCGTTCTCTCTTTGGTGGTCCAACTGGTGTGGTAAGACTGAGCTGCCAAG 358
 Qy 618 AlaleuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
 Db 359 GCCCTGGCTCAAGACGCTCTTTGATGATGAAATAATCTTTGGTGGAATTCACATGTCGAG 418
 Qy 638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHis 657
 Db 419 TATATGAGCAGCAGCATTCAGTTTCAAGATTGATTTGGGCTCTCTCTGGATATGTTGACAT 478
 Qy 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 Db 479 GATGAGGTGGCAACTTCACAGAGGAGTAAGAAGGAGGCTTATAGTGTGTACTGTTTC 538
 Qy 678 AspGluValGluIysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 Db 539 GATGAGTAGAAGACACATATCGCCGTATTCAATACCTCTCCAGGTTTTCAGCAT 598
 Qy 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
 Db 599 GGAAGGTAACTGATGGCCAGGCCGACCTGTAGACTTCACAAACACACATGATCATG 658
 Qy 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyIysValThrMetGlu 737
 Db 659 ACTTCAACCTTGGAGCTGAGCATCTCTCTCGGGTTTAGTAGGCAAGTGTACATGCAA 718
 Qy 738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
 Db 719 GATGCTGTACCCGGTGATGCGAGGAGTGAGNAGGCACCTTCAGGCCAGAGTTGCTTAAC 778
 Qy 758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
 Db 779 CGGCTGATGAATTTGGTATTGATCCCTCTCACATGATCACTGAGAAAGGTTGCC 838
 Qy 778 Arg 778
 Db 839 AGG 841

RESULT 4
 LOCUS CB894504

DEFINITION EST647296 HOGA Medicago truncatula cDNA clone HOGA-31M8, mRNA
 sequence.

ACCESSION CB894504

VERSION CB894504.1 GI:30101673

KEYWORDS EST.

SOURCE Medicago truncatula (barrel medic)

ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 846)
 Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
 Utterback,T., Cho,J. and Fraser,C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogallacturonides of DP 6-20
 Unpublished

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMDT76TK

More information is available at: www.medicago.org

Seq primer: Skmod (CTA GAA CTA gtg gat CC)

Location/Qualifiers

1. .846

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="Al7"

source

```

/db_xref="taxon:3880"
/clone="HOGA-31M8"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
BASE COUNT      270 a 139 c 243 g 194 t
UGIN

Alignment Scores:
ed. No.:      2,45e-110      Length:      846
ore:          1261.50      Matches:      249
Percent Similarity: 95.04%      Conservative: 19
st Local Similarity: 88.30%      Mismatches: 11
ery Match:      27.73%      Indels:      3
:              14      Gaps:      1

-09-812-350-17 (1-911) x CB994504 (1-846)
389 AspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSer 408
1 GACAGGCCATTGATTAGTTGATGGCATGCTTAATCTGAGAGTCACTTGATAT 60
409 GlnProGluGluLeuAspGlnLeuGluArgLysArgMetGlnLeuGluLeuHis 428
61 CAACCTGAGGAATTCACATCTTGAAGAGAGAATGCAATTAGAGTTGAATGTCAT 120
429 AlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuLeuGluValArgLysGlu 448
121 GCTCTAGAGAAAGAGAGGAGCAAGCTAGCAAGCCCGCTCTGTAGATGTACGAGAGAA 180
449 LeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGlu 468
181 CTTGATGACCTTGAGACAGACTTCAACCTCTGAAGATGAAGTATAGTAAGAGAAAGAG 240
469 ArgLysGluLeuLeuArgLysGlnLysGlnLysArgGluLeuMetPheSerLeuGln 488
241 AGGATTGATGAGATTCGAGGCTGAAGCAGAAACGTCGAGAGCTCCTCTTTCACATACAG 300
489 GluAlaGluArgTyrAspLeuAlaAlaGluAlaAspLeuArgTyrGlyAlaIleGln 508
301 GAGCAGAGAGGCGGTATGATCTAGCAGAGCTGCGGACCTGCGATATGTCGAATTGAA 360
509 GluValGluSerAlaIleAlaGlnLeuGluGlyThr-----SerSerGluGluAsn 525
361 GAGGTGAAACTGCAATTAAATAATCTTGAAGGTAGCACTGATGGGAACACGGATGAGAAC 420
526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThr 545
421 TTAATGTGTGCGAAACAGTTGACCGCGCAATAATAGCCGAGGTGTGTGCCGATGACT 480
546 GlyIleProValThrArgLeuGluGlnAsnGluLysGluArgLeuLeuGlyLeuAlaAsp 565
481 GGTATACAGTACTAGCTCGGCCCAAAATGAGAAAGCAAGGTGTGTGACTTGTGTGAC 540
566 ArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
541 AGACTGCACACTAGAGTCGTGGGACAGACCAAGCGGTTAATGCTGTGCTGAGGCTGTA 600
586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
601 TTGAGATCAAGAGCTGTTTGTAGGAAGACCCCAACCAACTGCTTCTTATTCCTT 660
606 GlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAsp 625

```

```

661 GGTCCGACTGCTGTTGGTAAACTGAGCTTGCAAGGCTCTTGACAGAGCACTATTGAT 720
626 AspGluAsnLeuLeuValArgLysMetSerGluTyrMetGluGlnHisSerValSer 645
721 GATGAAATCATCTGTTGGTGAATTCATGATGCGAATACATGGAACACACTCTGTATCA 780
646 ArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGlu 665
781 AGATTGATTGTCACCAACCAAGGCTATGTTGGACATGAGGAAGGGGTCAATTACAGAA 840
666 AlaVal 667
841 GCTGTA 846

RESULT 5
CB994478
LOCUS EST647270 HOGA Medicago truncatula cDNA clone HOGA-31G20, mRNA
DEFINITION sequence.
ACCESSION CB994478
VERSION CB994478.1 GI:30101647
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
REFERENCE 1 (bases 1 to 837)
AUTHORS Hahn,M.G., Ojansen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula treated with
oligogalacturonides of DP 6-20
JOURNAL Unpublished
COMMENT Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu
TIGR sequence name: MTMDT46TK
More information is available at: www.medicago.org
Seq primer: SKmod (CTA GAA CTA gTg gAT CC).
Location/Qualifiers
1..837
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Al7"
/db_xref="taxon:3880"
/clone="HOGA-31G20"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
BASE COUNT      258 a 147 c 236 g 196 t
ORIGIN

Alignment Scores:
Pred. No.:      2,47e-108      Length:      837
Score:          1240.50      Matches:      244
Percent Similarity: 94.24%      Conservative: 18
Best Local Similarity: 87.77%      Mismatches: 13

```



```

Query Match: 27.26% Indels: 3 Gaps: 1
09-812-350-17 (1-911) x CB894478 (1-837)

446 ArgLysLeuLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLys 465
|||||
3 CGAGAGAACTTGATGACTTGAGAGCAAGCTTCAACCTCTGAAGATGAAGTATAGTAAA 62

466 GluLysGluArgLysAspGluLeuLeuArgLysGlnLysArgGluGluLeuMetPhe 485
|||||
63 GAGAAAGAGAGGATTCATGAGATTCGAGGCTGAGCAGAAACGTCGAGAGCTCTCTTT 122

486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArgTyrGly 505
|||||
123 GCACATACAGAGGAGAGAGGCGGTATGATCTAGCAAGAGCTGCGGACCTCGATATGCT 182

506 AlaLeuGlnGluValGluSerAlaLeuAlaGlnLeuGluGlyThr-----SerSer 522
|||||
183 GCATTTGAAGAGGTGAAACTGCAATTAATAATCTTGAAGGTAGCACTGATGGGACACG 242

523 GluGluAsnValMetLeuThrGluAsnValGlyProGluHisAlaGluValValSer 542
|||||
243 CATGAGAACTTAATGTTGACGGAACAGTTGACCGGACCAATAGCGAGTTGTTAGC 302

543 ArgTyrThrGlyLeuProValThrArgLeuGlyGlnAsnGluLysGluGluLeuGly 562
|||||
303 CATGGACTGGTATACCACTGACTAGGCTCGGCCAAATAGAGAAAGCAAGGTTGGTTGA 362

563 LeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSer 582
|||||
363 CTGTGTGACAGACTGCACACTAGAGCTGTGGGACCAAGCAAGCGTTTAATGCTGTGCT 422

583 GluAlaLeuLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPhe 602
|||||
423 GAGGCTGTATTCAGATCAAGAGCTGTTTAGAGAGACCCCAACCACTGTTCTCTTC 482

603 LeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGln 622
|||||
483 TTATTCCTGTGTGCGACTGGTGTGGTAACTGAGCTTGCAAAAGGCTCTTGCAGAGCAG 542

623 LeuPheAspAspGluAsnLeuLeuValArgLysLeuMetSerGluTyrMetGluGlnHis 642
|||||
543 CTATTTGATGATGAATATCAGTTGTTGAGAAATTCATGTCGGAATACATGGAACAAC 602

643 SerValSerArgLeuLeuGlyAlaProProGlyTyrValGlyHisGluGluGlyGln 662
|||||
603 TCTGTATCAAGATTGATTGGTGACCAACCAAGGATGTTGACATGAGGAGGCGGTCAA 662

663 LeuThrGluAlaValArgArgProTyrCysValLeuLeuPheAspGluValGluLys 682
|||||
563 TTAACAGAGCTGTGAAGAGAGAGGACCATACAGTCTTGTACTCTTTGATGAAGTTGAGAAA 722

683 AlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAsp 702
|||||
723 GCACACATCTGTTATCAACACTCTTCTCCAGCTCTTGATGATGGAGGTTGACTGAC 782

703 GlyGlnGlyArgThrValAspPheArgAsnSerValLeuLeuMetThrSerAsn 720
|||||
783 GCACAGGCGAAGCTGCGATTTTGAAGAACTGTGATCATCATGACCTCTTAC 836

JUL 6
US
CB893989
INITIATION EST646781 HOGA Medicago truncatula cDNA clone HOGA-29N13, mRNA
sequence.
CB893989
CB893989.1 GI:30101158
EST.
MORDS
CE
MORGANISM
Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids.
; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 812)
Hahn, M.G., Ojansen-Reuhe, T., SamaC, D., Town, C.D., Van Aken, S.,
Utterback, T., Cho, J., and Fraser, C.M.
ESTs from roots of Medicago truncatula treated with
oligogalacturonides of DP 6-20
Unpublished
Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahncrcr.uga.edu
TIOR sequence name: MTWDM79TX
More information is available at: www.medicago.org
Seq primer: SKmod CTA GAA CTA gtg gat CC).
FEATURES
Location/Qualifiers
1..812
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Al7"
/db_xref="taxon:3880"
/clone="HOGA-29N13"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="X10LR"
/clone_lib="HOGA"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
BASE COUNT 258 a 137 c 233 g 184 t
ORIGIN
Alignment Scores: 2.25e-103 Length: 812
Pred. No.: 1188.50 Matches: 233
Percent Similarity: 94.07% Conservative: 21
Best Local Similarity: 86.30% Mismatches: 13
Query Match: 26.12% Indels: 3
DB: 14 Gaps: 1
US-09-812-350-17 (1-911) x CB893989 (1-812)
QY 420 ArgMetGlnLeuGluLeuGluLeuHisAlaLeuGluArgGluLysAspLysSerLys 439
|||||
Db 1 AGAATGCAATTAAGATTTGAATGCTGCTAGAGAAAGAGAGCAAAAGCTAGCAA 60
QY 440 AlaArgLeuLeuGluValArgLysGluLeuAspLeuArgAspLysLeuGlnProLeu 459
|||||
Db 61 GCCCGCTCTGTAGATGTACGAGAGAACTTGTGATCTTGAGAGCAAGCTTCAACCTCTG 120
QY 460 ThrMetLysTyrArgLysGluLysGluArgLysGluLeuAspGluLeuGlnLys 479
|||||
Db 121 AAGATGAAGTATAGTAAAGAGAGAGAGAGAGTTCATGAGATTCGAGGCTGAGCAGAAA 180
QY 480 ArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAla 499
|||||
Db 181 CGTGAAGAGCTCTCTCTTTCGACTCAGAGAGGCGGCTGATGATCTAGCAAGAGCT 240
QY 500 AlaAspLeuArgTyrGlyAlaLeuGlnValGluSerAlaLeuAlaGlnLeuGluGly 519
|||||
Db 241 GCGGACCTGCGATATGTTGCAATTTGAAGAGGTGAAACTGCAATTAATAATCTTGAAGGT 300
QY 520 Thr-----SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
|||||

```

301 AGCACTGATGGGACACCGATGAGCACTTATGTTGACGGAAACAGTTGGACCGACAA 360
 537 IleAlaGluValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
 361 ATAGCCGAGGTGTTAGCCGATGAGCTGTTATACCACTGACTAGCTCGCCAAATGAG 420
 557 LysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
 421 AAAGCAAGGTGGTGGGACTTGGTGACAGACTGACACTAGAGTCGTGGGACAGACAA 480
 577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyValGln 596
 481 GCGGTATATGCTGTTGCTGAGCTGTTATGATGATGAGCTGTTGAGGAGACCCAA 540
 597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
 541 CAACCACTGCTGCTCTTATTCCTTGGTCCGACTGCTGTTGTTGTAACACTGAGCTTGA 600
 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValAlaGlyIleAspMetSer 636
 601 AAGCTCTTTCAGAGCGAGCTATTTGATGATGATAAATCAGTTGGTGAATTTGACATGTCG 660
 637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGly 656
 661 GAATACATGGAACACACTCTGTATCAGATTGATTGTCACACACAGGGTATTTGGA 720
 657 HisGluGluGlyGlnLeuThrGluAlaValArgArgProGlyCysValIleLeu 676
 721 CATGAGGAAGGGGTCAATTAACAGAGCTGTAGGAGAGGCGCATACAGTGTGTACTC 780
 677 PheAspGluValGluLysAlaHisValAla 686
 781 TTGATGAAGTTGAGAAAGCACACATCT 810

SULT 7
 648761
 CUS
 EST510380 HOGA Medicago truncatula cDNA clone phOGA-23B24 5' end,
 mRNA sequence.
 BG648761
 BG648761.1 GI:13783873
 EST.
 Medicago truncatula (barrel medic)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 818)
 Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
 Utterback, T., Cho, J. and Fraser, C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogalacturonides of DP 6-20
 Unpublished
 Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu
 G392367e TIGR sequence name: MTWCP12TK More information is
 available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gtg GAT CC).
 Location/Qualifiers
 1. .818
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="Al7"
 /db_xref="taxon:3880"
 /clone="phOGA-23B24"
 /tissue type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C

with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_host="XLOLR"
 /clone_lib="HOGA"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-zap phage using Ex-assist
 helper phage and propagated in SOUR cells."

BASE COUNT 239 a 151 c 230 g 198 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,27e-103 Length: 818
 Score: 1188.50 Matches: 232
 Percent Similarity: 93.75% Conservative: 23
 Best Local Similarity: 85.23% Mismatches: 14
 Query Match: 26.12% Indels: 3
 DB: 10 Gaps: 1

US-09-812-350-17 (1-911) x BG648761 (1-818)

QY 484 MetPheSerLeuGlnGluAlaGluArgTrpAspLeuAlaArgAlaAspLeuArg 503
 Db 2 CTCCTTGCACCTACAGAGGCGAGAGGCGGTATGATCTAGCAGAGCTGCGGACTGCGA 61
 QY 504 TyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr----- 520
 Db 62 TATGTCGAATTTGAAGAGGTGGAACCTCAATTTAAATCTTTGAAGGTAGCCTCATGGG 121
 QY 521 SerSerGluGluAenValMetLeuThrGluAenValGlyProGluHisIleAlaGluVal 540
 Db 122 AACCGGATGAGACTTAATGTTGCGGAAACAGTTGGACCGGCAAAATAGCGAGGTT 181
 QY 541 ValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
 Db 182 GTTACCCGATGAGCTGTTATACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 241
 QY 561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
 Db 242 GTTGGACTGTTGTCGACAGACTGCACACTAGAGTCTGGGACAGACCAAGCGGTTAATGCT 301
 QY 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 Db 302 GTTGCTGAGGCTGTATTGAGATCAAGAGCTGGTTAGAGAGCCCAACCAACCACTGGT 361
 QY 601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
 Db 362 TCCTTCTTATTCCTTGGTCCGACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 421
 QY 621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGlu 640
 Db 422 GAGCAGCTATTGTTGATGAAATCACTGTTGAGAAATTCACATGTCGGAATACATGGAA 481
 QY 641 GlnHisSerValSerArgLeuIleGlyValaProProGlyTyrValGlyHisGluGluGly 660
 Db 482 CAACACTCTGTATCAAGATTGATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 541
 QY 661 GlyGlnLeuThrGluAlaValaArgArgProGlyCysValIleLeuPheAspGluVal 680
 Db 542 GGTCAATTAACAGAGCTGTAGGAGAGGCGCATACAGTGTGTACTCTTTGATGAAGTT 601
 QY 681 GluLysAlaHisValAlaValaPheAsnThrLeuGlnValLeuAspGlyArgLeu 700
 Db 602 GAGAAACACACATCTGTATTCAACACTCTTCTCAAGTCTTGGATGATGGAGGTTG 661
 QY 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
 Db 662 ACTGACGACACAGGAGACTGTGGATTGTTAGAAACACTGTGATCATGATGACCTTAAC 721
 QY 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740

ATURES
 source


```

/cultivar="A17"
/db_xref="taxon:3880"
/clone="HOXA-28N14"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_hosts="XLOLR"
/clone_lib="HOXA"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using EX-assist
helper phage and propagated in SOLR cells."
SE COUNT 233 a 149 c 226 g 196 t
CIGGIN

Alignment Scores:
ed. No.: 1,46e-101 Length: 804
ore: 1169.50 Matches: 230
cent Similarity: 93.63% Conservative: 20
st Local Similarity: 86.14% Mismatches: 14
Match: 25.70% Indels: 3
: 14 Gaps: 1

-09-812-350-17 (1-911) x CB893763 (1-804)
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaAla 500
2 GAAGAGCTCTCTTTCACCTACAGGAGCGAGAGCGGTATGATCTAGCAGAGCTGG 61
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluThr 520
62 GACCTCGATATGTCATTCATTAAGAGGTGGAACTGCAATTAATAATCTTGAAGTAGC 121
521 -----SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
122 ACTGATGGGAACACGGATGAGACTTAATGTTGACGGAAACAGTTGGACCGGACCAATA 181
538 AlaGluValValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
182 GCGAGGTGTTAGCGGATGAGCTGTATACAGTACTAGCTCGGCCCAAAATGAGAAA 241
558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
242 GCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 301
578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
302 GTTATGCTGTGCTGAGCGCTGATTGAGATCAGAGCTGGTTTGGGAAGACCCCAACA 361
598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
362 CCAACTGGTTCCTCTTATCTCTGTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 421
618 AlaLeuAlaGluGlnPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
422 GCTCTTCGAGGACGACTATTGATGATGATAAATCAGTTGGTGGAGAAATGACATGTCGAA 481
638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis 657
482 TACATGGGAACAACACTCTGTATCAAGATTGATTGTTGTCACCAACAGGATGTTGGACAT 541
658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
542 GAGGAAGCGGGTCAATTAAAGAGCTGTAAGGAGAGGCCATACAGTGTTGTACTCTTT 601
678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
602 GATGAGTTGAGAAAGCACACATCTGTATTCAACACTCTTCTCCAAGTCTTGGATGAT 661

```

```

QY 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMet 717
Db 662 GGGAGTGTGACTGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 721
QY 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
Db 722 ACCTCTAACCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 781
QY 738 ValAlaArgAspCysValMet 744
Db 782 GCTGCTCGGATCGAGTGATG 802

BQ839239 740 bp mRNA linear EST 08-AUG-2002
WHB4163_H06_O112S Wheat CS whole plant cDNA library Triticum
aestivum cDNA clone WHB4163_H06_O11, mRNA sequence.
BQ839239
BQ839239.1 GI:22143561
EST.
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 740)
Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J.,
Lazo, G.R., Rausch, C.J., Wilson, C. and Woo, J.
The structure and function of the expressed portion of the wheat
Genomes - Chinese Spring whole plant cDNA library
Unpublished
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
Location/Qualifiers
1..740
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHB4163_H06_O11"
/tissue_type="Roots, leaves, crown, stem and sheath"
/dev_stage="Adult"
/lab_host="E. coli SOLR"
/clone_lib="Wheat CS whole plant cDNA library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid
pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant
tissues from wheat cv. CS grown to full tillering stage in
greenhouse were collected at University of California,
Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
RNA was prepared from leaves (young leaf and third leaf),
whole roots, crown, stem and sheath tissues, and then
equal quantities of RNA were pooled from the these
samples. PolyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo excised
to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.
Akhunov, J. Dvorak) at the University of California,
Davis. Colony plating, plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
BASE COUNT 136 a 212 c 282 g 110 t
ORIGIN
Alignment Scores: 6.23e-100 Length: 740
Pred. No.: 1152.00 Matches: 226
Score:

```

```

Percent Similarity: 98.35%      Conservative: 13
Local Similarity: 93.00%      Mismatches: 4
Every Match: 25.32%          Indels: 0
: 13                          Gaps: 0

```

-09-812-350-17 (1-911) x BQ839239 (1-740)

153 GluSerAlaSerGlyAspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuVal 172
10 GAGTCCGCGTGGGGGACACCACTTCAGCGCCTCAAGACGTACGGCGCGACCTCGTG 69
173 GluGlnAlaGlyLysLeuAspProValIleGlyArgAspGluLeuLeuLeuVal 192
70 GAGGTGGCGGGCAAGCTGGACCCGGTTCATCGGCGCGACGAGAGATCCGGCGCGTGGTG 129
193 ArgIleLeuSerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGly 212
130 CGGATCTGTGCGGGCGACCAAGAACAACCCCGTCTGATCGGCGAGCCCGCGGTGGCG 189
213 LysThrAlaValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSer 232
190 AAGACCGCGTGTGGAGGGGCTCGCGCAGCGCGTCTGTCGGCGGCGACGTCCCGCAGCAAC 249
233 LeuThrAspValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyr 252
250 CTCCTGGACGTGGCGCTGGTTCGGCTGGACATGGGCGCGCTCTGGTCCGCGCCCAAGTAC 309
253 ArgGlyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGly 272
310 CGCGCGAGTTCGAGGAGCGGCTCAAGGCGGTGCTCAGAGAGTGGAGGAGCCCGAGGGG 369
273 LysValIleLeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGly 292
370 AAGGTGATACTGTTCATCAGCAGATACACCTGGTGTCTCGGCGCGCGGCGAGCGAGGGG 429
293 SerMetAspAlaAlaasnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIle 312
430 TCCATGGACGCGCGCAACCTGTTCAAGCGATGCTGGCGAGGGGCGAGCTCAGGTGCATT 489
313 GlyAlaThrThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArg 332
490 GCGCGCAGCACCTCGAGGAGTACAGGAAGTACGTGGAGAGGACGCGCGCTTCGAGAGG 549
333 ArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArg 352
550 CGGTTCCAGCAGGTGTCTGTGGCAGAGCCACGCTCCCGCACCGTCCAGCATCTGNGA 609
353 GlyLeuLysGluLysTyrGluGlyHisGlyValAlaGlyIleGlnAspArgAlaLeuIle 372
610 GGGCTCAGGGAAGATACGAGGGGCACCATGTTGTGAGGATTCAGGACCGCGCGTCTGTC 669
373 AsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIle 392
670 ATTGCCGCGCAGCTCTCGTCGAGGTACATCATGGGTGCTCATCTGCCCTGATATAAGCAATC 729
393 AspLeuVal 395
730 GATTGGTT 738

U/LT 11

10803

US

INITIAL

SESSION

NOTES

WORDS

PRICE
CONTAINING

ORGANISMS

REFERENCE

AUTHORS	TITLE	JOURNAL	COMMENT
---------	-------	---------	---------

Uratsu, S. L., Fisk, H. J. and Dandekar, A. M.
Analysis of peel specific genes in Citrus
Unpublished
Contact: Abhaya Dandekar, PhD
CAES Genome Facility
UC Davis, Department of Pomology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 752 7784
Fax: 530 752 8502
Email: amandandekar@ucdavis.edu
Seq primer: TC CGAATCTGCAGCAGC.

FEATURES

```

1. 1.722
/organism="Citrus sinensis"
/mol_type="mRNA"
/cultivar="Washington navel orange"
/db_xref="taxon:2711"
/clones="ALBEDO0002_IVAF_D06"
/sex="Hermaphrodite"
/cell_type="albedo"
/dev_stage="Mature fruit sample - c
/lab hosts="Xil10-Gold"
/clone lib="Mature Albedo"
/note="Organ: Peel/rind; Vector: pT
Site 2: 36A1; Mature citrus fruits
trees growing in the 'Citrus variety
Wolfskill experimental orchard local
California (USA). Fruit was collected
between 1 and 3 PM and stored at 4°C
separated from the rest of the peel.
RNA using Trizol reagent from INVITRO
was constructed using the SMART cDNA
). The amplified library was an mRNA
DNA containing the cDNA library was
resultant bacterial population. plasm
transformed into ultra competent E.
Stratagene). Transformants were pla
(2000 cfu/tray), picked using a Qbo
well dishes".

```

BASE COUNT	202 a	114 c	213 g	189 t	4 others
ORIGIN					

Alignment Scores:

Pred. No.:	2,096-96	Length:	722
Score:	1115.00	Matches:	220
Percent Similarity:	96.6%	Conservative:	11
Best Local Similarity:	92.05%	Mismatches:	8
Query Match:	24.51%	Indels:	0
DB:	14	Gaps:	0

US-09-812-350-17 (1-911) x CB610803 (1-722)

186	QY	GlulGlulleArgA-gValValArgilleuSerArgThrLyAsnAanProValLeu	205
4	Db	GAAGAATTTAGAACGTTTGTTCGAATTCGTCAAGGAGGACTAAAAACAATCTGCTCTT	63
206	QY	lleglyGlupProGlyValGlyLySThrAlaValValGluGlyLeuAlaGlnArgilleVal	225
64	Db	ATTGAGAGCCAGAGATTGGTAAACCGCTGTGGTGAAGAGATTGCTCTAGAGAATTGTG	123
226	QY	LySglyAspValProAanSerLeuThrAspValArgLeuLeuSerLeuAspMetGlyAla	245
124	Db	CGAGAGATGGCTTCACCAACCTTCGTGATGTGAGGCTCATTTGCTTTTGATATGGTGCT	183
246	QY	LeuValAlaGlyAlaLySThrArgGlyGluPheGluArgLeuLySserValLeuLyS	265
184	Db	CTGCTTGAGCGCCCAAGTATAGGCGAGAGTTTGAAGAGAGGTTGAAGGCTGTTCTGAAA	243
266	QY	GluValGluAspAlaGluGlyLySValIleLeuPheIleAspGluIleHisLeuValLeu	285
244	Db	GAGTAGAAGAGCTCAGGAGAAAGTGATTTGTTTATTGATGAGATTACACTTGTGCTT	303
286	QY	GlyAlaGlyLySThrGluGlySerMetAspAlaAaAsnLeuPheLySProMetLeuAla	305

P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L., and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compgenomics.ucdavis.edu/>
Unpublished
Contact: Alexander Kozik [R.W. Michelmore]
Department of Vegetable Crops, R.W. Michelmore Lab
University of California at Davis (UCD)
Amundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Cont194657, see <http://cgdb.ucdavis.edu/>
for details.
Plate: QGF5 row: P column: 11.
Location/Qualifiers
1. .727
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L. serriola"
/db_xref="taxon:4236"
/clone="QGF5P11"
/lab_host="E.coli"
/note="Vector: pBRCN5f1AB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformation made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgdb.ucdavis.edu/>
TAG LIB=QG_EFGHJ lettuce serriola
TAG TISSUE=flowers post-fertilized
TAG_SEQ=RGCCATCGG"

FEATURES

source

206 a 110 c 210 g 200 t 1 others

Segment Scores:

ad. No.: 2,36e-96 Length: 727
ore: 1114.50 Matches: 222
cent Similarity: 97.11% Conservative: 13
t Local Similarity: 91.74% Mismatches: 6
ry Match: 24.49% Indels: 1
13 Gaps: 1

09-812-350-17 (1-911) x BQ993972 (1-727)

147 LysGluGlyLysLysValGluSerAlaSerGlyAspThrAsnPhleGlnAlaLeuLysThr 166
3 AAGAAGGGAAGAGGTGAAGCGCTTCAGGGGATCTAATTTCCAGCTCTCAAGACA 62
167 TyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspProValIleGlyArgAspGlu 186
63 TATGTCGTGATCTTGTAGAGCAAGCGGAGCTTGATCTCTGTAATTTGGAGAGATGAG 122
187 GluLeuArgArgValValArgIleLeuSerArgArgThrLysAsnAsnProValIleLeu 206
123 GAAATTAGCAGATGATAGGATTTTATCAGAGAGGACTAGAACATCTCTGTTCTTATT 182
207 GlyGluProGlyValGlyLysThrAlaValValGluGlyLeuAlaGlnArgIleValLys 226
183 GGTGAACCTGTTGGGAAAACTGCTGTTCTGAAGGATAGCTCAAGGATCTTAGA 242
227 GlyAspValProAsnSerLeuThrAspValArgLeuLeuSerLeuAspMetGlyAlaLeu 246
243 GGAGATGTCCTCCAGTAATCTGCTGATGTAGAGGTATCGCATTTGGATATGGAGCTTTA 302
247 ValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeuLysSerValLeuLysGlu 266
303 ATTGCTGGAGCAAGTATAGAGGGGAGTTTGGAGAGAGACTGAAGCAGTTTGAAGAA 362

QY 267 valGluAspAlaGluGlyLysValIleLeuPheLeuAspGluIleHisLeuValLeuGly 286
Db 363 GTGGAAGACGCTGAAGGGAAGGTGATATTATTCAATGATGAATCCACCTGTTCTTGA 422
QY 287 AlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArg 306
Db 423 GCGGTCGACAGAGAGGATCCATGATGCTGCAATCTGTTCAAGCCCATGCTTCTAGG 482
QY 307 GlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyrArgLysTyrValGluLys 326
Db 483 GGTCAACTCCGTCGATTTGGTCAACTACTTTTGAAGAGTATAGGAATATGTGAAAAA 542
QY 327 AspAlaAlaPheGluArgArgPheGlnGlnValTyrValAlaGluProSerValProAsp 346
Db 543 GATGCTGCTTTTGAAGACGTTTCCAGCAGGTTTGGTGGCTGAACCAAGTGTCTCTGAC 602
QY 347 ThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGlyHisGlyValArgIle 366
Db 603 ACCATAAGCATCTTAGAGGGTTGAAAGAGATATGAAGT---CATGGTGTAGGATT 659
QY 367 GlnAspArgAlaLeuIleAsnAlaGlnLeuSerAlaArgTyrIleThrGlyArgHis 386
Db 660 CTTCATGCTGCTCTTGTGTTGCTGCTCACTCAAGCCGATACATACTGCACGTTT 719
QY 387 LeuPro 388
Db 720 CTGCT 725

RESULT 14

BG648385

LOCUS

DEFINITION

EST510004 HOGA Medicago truncatula cDNA clone PHOGA-19L1 5' end,

mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .736
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="PHOGA-19L1"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C with 0.5 mg/ml oligogalacturonides (DP 6-20) in the presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Unizap XR vector from
Stratagene and packaged using GigaPack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."

201 a 88 c 219 g 228 t

Alignment Scores:

5.18e-96 Length: 736
1111.00 Matches: 224
96.73% Conservativity: 13
91.43% Mismatches: 8
24.42% Indels: 1
10 Gaps: 0

-09-812-350-17 (1-911) x BG648385 (1-736)

```

97 GlnAlaAlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGly 116
|||||
1 CAAGCTGCTCAGAAATTA-CGTGGTGATCTATTGATGCTGATGCTGATGCTGCTGGA 59
|||||
117 LeuLeuGluAspSerGlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArg 136
|||||
60 ATTCTTGAGGATCTCAGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 119
|||||
137 ValLysSerGluValGluLysLeuArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 156
|||||
120 GTTAAAGCTGAGGTGAAAGAGCTGAGGGTAAAGAGTGAAGAGTGAAGAGTGAAGAG 179
|||||
157 GlyAspThrAsnGlnAlaLeuLysThrTyArgAspLeuValGlyValAlaGly 176
|||||
180 GGTGATACGAAATTTTCAAGCATTTGAAGCATTTGATGATGATGATGATGATGATGAT 239
|||||
177 LysLeuAspProValIleGlyArgAspGluLysLeuArgValValAlaGlyLeuSer 196
|||||
240 AAGCTTGATCTGTTATTTGACGCGAGAGAGATGCAAGAGATGCAAGAGATTTGTCA 299
|||||
197 ArgArgThrLysAsnAsnProValLeuLysGlyGlyGlyGlyGlyGlyGlyGlyGly 216
|||||
300 AGGAGGAGTAAAGATATATCCAGTCTTATTGTTGAGCTGCTGCTGCTGCTGCTGCT 359
|||||
217 ValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 236
|||||
360 GTTCAAGGGTGGCTCAGAGATTTGAGAGGATGATGCTGATGATGATGATGATGATG 419
|||||
237 ArgLeuLysSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGlu 256
|||||
420 AGGTTAATCTCTGGATATGGGAGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479
|||||
257 GluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyValIleLeu 276
|||||
480 GAAGAGAGGTTGAAGCTGTTTGAAGAGATTTGAAGAGATTTGAAGAGATTTATCTT 539
|||||
277 PheLeuAspGluLeuHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 296
|||||
540 TTCAATGATGAGATTTCACTGTTCTTGGAGCTGGTAGACAGAGATCAATGGATGCT 599
|||||
297 AlaAsnLeuPheLysProMetLeuAlaArgGlyGlyGlyGlyGlyGlyGlyGlyGly 316
|||||
600 GCTAATCTTTTAAAGCAATGCTTCTGCTGAGAGCTGCTGATGCTGCTGCTGCTGCT 659
|||||
317 LeuGluGluThrArgLysThrValGluLysAspAlaAlaPheGluArgPheGlnGln 336
|||||
660 CTTGAGAGATAGGAGATGATGTTGAAAGAGATGAGCATTTGAAAGAGATTTTCAACAG 719
|||||
337 ValTyrValAlaGlu 341
|||||
720 GTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 734

```

BQ862449 729 bp mRNA linear EST 14-AUG-2002

SUBMIT 15
862449
CUCS

DEFINITION

QGC21B24.YG.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone
QGC21B24, mRNA sequence.
BQ862449
VERSION
BQ862449.1 GI:22247914
KEYWORDS
EST.
SOURCE
Lactuca sativa
ORGANISM
Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.

REFERENCE

1 (bases 1 to 729)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.

TITLE

Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/

JOURNAL

Unpublished

COMMENT

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
Belongs to contig QG_CA_Contig4657, see http://cgpdb.ucdavis.edu/
for details.

Plate: QGC21 row: B column: 24.

FEATURES

location/Qualifiers

1..729

/organism="Lactuca sativa"

/mol_type="mRNA"

/cultivar="Salinas"

/db_xref="taxon:4236"

/clone="QGC21B24"

/lab_host="E.coli"

/clone_lib="QG ABCDI lettuce salinas"

/note="Vector: pBRCNMFAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB-QG ABCDI lettuce salinas
TAG_TISSUE=shoots environmental stress
TAG_SEQ=TCGCACGGG"

BASE COUNT 209 a 127 c 238 g 155 t

ORIGIN

Alignment Scores:

Pred. No.: 8.04e-95 Length: 729

Score: 1098.50 Matches: 215

Percent Similarity: 96.71% Conservativity: 20

Best Local Similarity: 88.48% Mismatches: 7

Query Match: 24.14% Indels: 1

DB: 13 Gaps: 1

US-09-812-350-17 (1-911) x BQ862449 (1-729)

QY 466 GluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeuMetPhe 485

DB 3 GAAAGAGAGAGATTTGATGAAATCCGAGACTGAACAGAAAGAGAGAGCTTTGGTG 62

QY 486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaAlaArgAlaAlaArgTyrGly 505

DB 63 GCATCTACAGAGAGAGAGAGAGAGATGATGATTTAGCAGAGCTGCAGATTTGAATTCGA 122

QY 506 AlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsn 525

123 GCAGTTCAGAACTGGAAACAGCAATAGCAAACTCGAAGGAACA---ACAGATGAAAT 179
526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTrpThr 545
180 GTGATGTTGACAGAAACAGTGGACCCGATCAGATAGCTCAGGTGGTGGTGGATGACC 239
546 GlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAsp 565
240 GGGATCCCGGTGACTCGACTTGGAAACCAATGAGAAAGAGCGGTGGTGGTGGTGGT 299
566 ArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
300 CGGTGTCATCAACGGGTCGGGTCAAGACGCGGGTGAACGCGGTGGTGGTGGTGGT 359
586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
360 CTGAGGTCACGAGCGGGCTCGGAAGGGCTCAGCAGCCACCGGATCCTTCCTGTTCTG 419
606 GlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAsp 625
420 GGTCCCAACAGGTGTTGGAAAAACAGAGCTTGCAGAGGCTCTTCAGAGCAGTTGTTGAT 479
626 AspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSer 645
480 GATGAGAAAGCTGATGATGAAATCGATATCGAGTATATGAGAGCAGCATTCGGTGCA 539
646 ArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGlu 665
540 CGGCTGATCGAGCTCCGCCAGGGTATGTTGGTCAAGAGCGGGCAGTTGACCGAG 599
666 AlaValArgArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisVal 685
600 GCTGTGAGAGGAGACCATACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 659
686 AlaValPheAsnThrLeuGlnValLeuAspAspGlyValArgLeuThrAspGlyGlnGly 705
660 TCGGTGTTCAATACGCTTCTTCAGATGTTGATGAGAGGTTGACCGGATGGTCAAGGT 719
706 ArgThrVal 708
720 CGAACGGTT 728

Search completed: February 13, 2004, 04:58:38
time : 3857 secs

